Fringe Analysis

Yves Surrel

National Institute of Metrology, CNAM/INM,
292, rue St.Martin, F-75141 Paris Cedex 03, France
surrel@cnam.fr

Abstract. Fringe analysis is the process of extracting quantitative measurement data from fringe – or line – patterns. It usually consists of phase detection and phase unwrapping. Phase detection is the calculation of the fringes phase from the recorded intensity patterns, and this issue represents the major part of the material in this review. Different techniques for this phase calculation are presented, with special emphasis on the characteristic polynomial method, allows us permits to easily design customized algorithms coping with many error sources. For reference, a table presenting the properties of almost all algorithms which have been in recent years is provided in the Appendix. A generic method allowing us to quantitatively evaluate the phase errors and the effect of noise is presented here for the first time. Finally, some elements regarding the complex problem of phase unwrapping are given.

1 Introduction

Most optical methods for metrology require the processing of a fringe (or line) pattern. The intensity pattern which is provided by optical methods encodes the physical quantity which is measured, and this intensity varies as the cosine of a phase which is most often directly proportional to that physical quantity. Fringe processing is the extraction of the phase field from one or many intensity fields which have been acquired. The process of phase detection has received considerable attention during the last decade, as video acquisition and computer processing became more powerful and less expensive.

A typical example of a fringe pattern is presented in Fig. 1a. This example is a Michelson-type interferogram, and the fringe phase is directly proportional to the deviation of the measured surface from a reference plane. The problem is to extract the corresponding phase field represented in Fig. 1b, so that the profile of the surface can be obtained. The real fringe pattern exhibits usually variable bias and contrast, as can be seen when comparing Figs. 1a,c, and the phase extraction strategy has to take this into account. Finally, the phase which is obtained within the interval $[-\pi, +\pi]$ has to be unwrapped as in Fig. 1d to faithfully map the real profile. Note that very often the phase is only meaningful up to an additive constant.

This process should be compared to the historical approach using fringe orders, when fringes had to be counted manually throughout the field to

P. K. Rastogi (Ed.): Photomechanics, Topics Appl. Phys. **77**, 55–102 (2000)

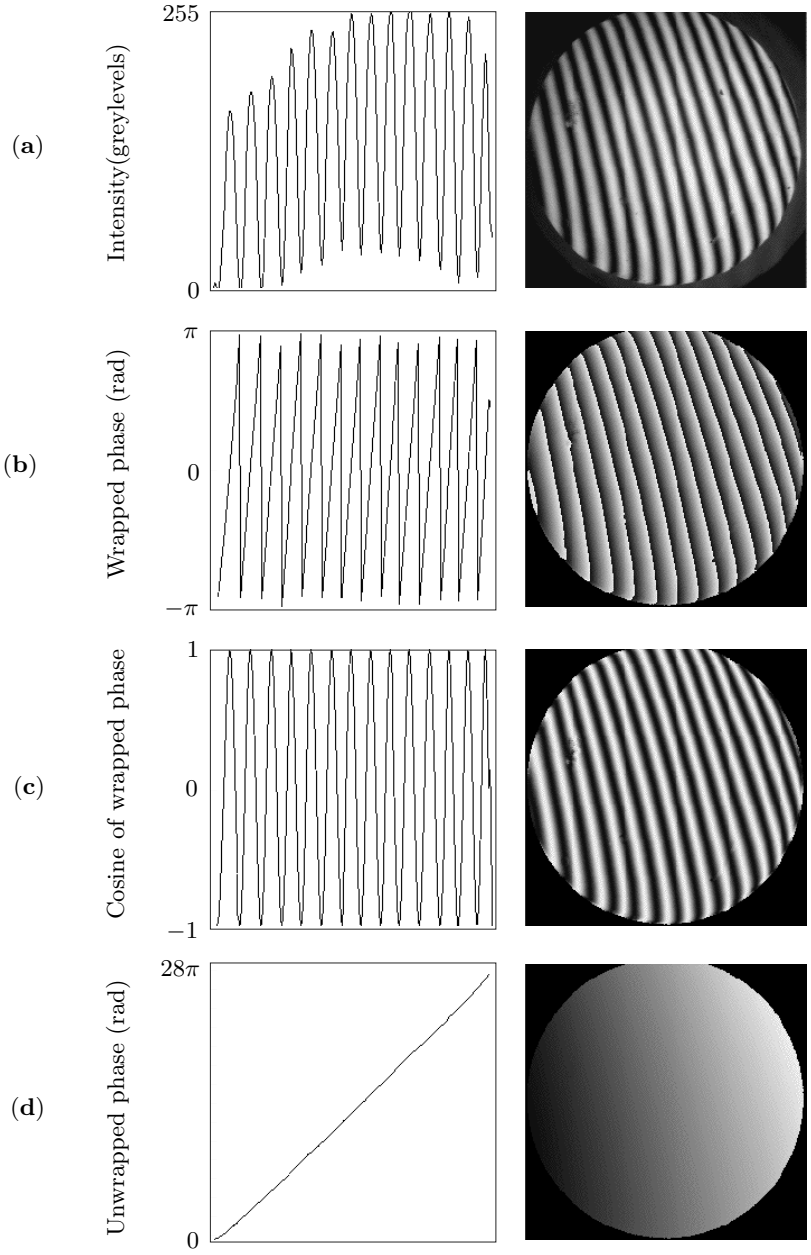


Fig. 1. Fringe processing (in all figures, the *left-hand side* displays a horizontal cross section across the center of the field): (**a**) fringe pattern, (**b**) corresponding wrapped phase, (**c**) ideal fringes, obtained from the cosine of the wrapped phase, and (**d**) unwrapped phase

make a quantitative measurement. The fringe order is the integer part of the division of the unwrapped phase by 2π. Actually, modern phase detection methods provide at any pixel not only the integer fringe order (once the phase is unwrapped) but also its fractional part, which can be written $\phi/2\pi$, where ϕ is the wrapped phase.

An important issue is the assessment of the relation between the phase and the physical quantity to be measured, when this relation is not linear. This is of particular importance in profilometry when using the fringe projection technique. However, this last topic is beyond the scope of this review.

2 Phase Evaluation

In this section, the problem of calculating a phase map from one or many fringe patterns is addressed. The phase detection process is first presented, and the discussion is centered on the role and properties of the detection "algorithm". The characteristic polynomial (CP) theory is used, as it provides on this topic a comprehensive set of results and requires only very simple calculations. Then the two major issues of phase error evaluation and noise effect are discussed.

2.1 Local and Global Strategies

Today, two main tracks can be followed for phase detection, which can be called "global" and "local" approaches.

- In the global approach, often called the Fourier transform technique [1,2,3,4,5,6,7,8], the whole intensity field is processed at the same time. The output phase data depend on the whole set of data in the intensity field. This technique cannot always be used: a carrier is necessary so that the phase modulation can be identified in the Fourier domain as the broadening of a well-defined frequency peak. This technique has two main advantages: first, only one intensity field is necessary, i.e. one image has to be grabbed, and dynamic events can be studied, and second, large phase modulations can be detected. This approach is presented at the end of this chapter. Using the vocabulary of signal processing, the procedure corresponds to extracting the 'analytic signal' from the intensity field. For metrological applications, the main disadvantage of this technique is that the error on the calculated phase is difficult to evaluate, especially near edges. Another disadvantage is that it is not easy to fully automate, as some filtering in the Fourier plane has to be done, that may depend on the pattern to be analyzed.
- The local approach is called phase-stepping or phase-shifting, and the phase is calculated from local intensity information. Two procedures exist, namely temporal and spatial phase-stepping. In temporal phase-stepping [9,10,11,12,13,14], a number of images are taken with a constant

and known phase shift δ introduced between successive images. There is no requirement on the fringe pattern, and the calculated phase at a given pixel depends only on the intensities recorded at the same pixel, unless spatial smoothing is used. Obviously, dynamic events cannot be studied as the phase-stepping takes some time. In spatial phase-stepping [15,16,17], a carrier is necessary so that a given number N of contiguous pixels sample a period of the intensity signal ("tuning" requirement). In this case, the values recorded by adjacent pixels play the role of the phase-stepped intensities and the phase can be calculated from one single frame. However, this tuning requirement is quite stringent and the method is often not applicable. It is especially suited to the grid method used for measuring displacements.

We can also mention an approach using wavelet analysis, which is also a kind of local detection. It will not be presented here, but a rather detailed analysis can be found in [18]. For completeness, a recent work concerning the use of adaptive quadrature filters must be cited [19].

In this review, we focus on the second (local) approach. The formula which is used to calculate the phase from the recorded intensities is usually called the 'algorithm'. Many algorithms have been proposed in the literature, and it is necessary to understand what are their respective advantages and disadvantages. In this chapter, we shall use the term 'phase-stepping' rather than 'phase-shifting'. The difference is that the shifter stops during image grabbing in the first case and not in the second one, thus decreasing the contrast. However, it is worthwhile mentioning that phase-shifting performs some amount of time averaging. This corresponds to a low-pass filtering which may reject part of the unwanted high frequency phase perturbations due to parasitic mechanical vibrations or air turbulence. The global approach will be shortly addressed in Sect. 2.3.

2.2 Local Phase detection

There are a number of systematic errors which can be introduced in the phase-stepping/phase detection process [20,21,22,23]. The main ones are listed below.

- There are many cases where the periodic intensity signal which encodes the physical quantity to measure is not sinusoidal. Many examples can be given. Moiré fringes have a profile which is rather triangular. In grid methods, the signal may be close to a binary one when black and white lines are used. Also, the Ronchi test provides fringes which have no reason to be sinusoidal, as they are the shadow of the Ronchi ruling placed in the beam path. In interferometry, some amount of multiple reflection may also lead to fringes which do not have an exact sine profile. Another very important example is when the detector exhibits some amount of

nonlinearity, in which case a perfect input sine is distorted and harmonics are generated.

- In case of a phase-step miscalibration, the phase variation between two images is not exactly equal to its nominal value δ. Two typical examples are: in interferometric techniques, a miscalibration of the piezoelectric transducer (PZT) which is used to move a mirror or rotate a parallel plate to vary the reference path length, and in the grid technique, the analysis of the strained state of a sample using spatial phase-stepping. There are also cases where the miscalibration is not homogeneous across the field of view. This happens, for example, when a phase-stepping mirror exhibits some amount of parasitic tilting during its piston movement.
- The bias variation is encountered in interferometry when the source is a laser diode, the intensity of which varies with time. It is also common in moiré and grid techniques when the field illumination is nonuniform.
- Vibrations can also cause systematic errors, for example when they make the optical path difference vary sinusoidally during temporal phase-stepping in interferometry. Indeed in that case, the electromagnetic phase variation due to the vibration is correlated with the sampling frequency.

For users, the question is to choose or design the algorithm which will cancel the effects of one or many of the above causes of errors, depending on the particular setup. Recently, several methods have been proposed for designing customized algorithms: the Fourier transform (FT) method [24], the characteristic polynomial (CP) method [25], the data-averaging/data-windowing method [26,27,28] and the recursion method [29]. It is shown in the following how these different approaches can be related. From a practical point of view, the CP approach requires very simple calculations, and provides a comprehensive set of results.

2.2.1 Characteristic Polynomial and Diagram

The intensity field which is recorded during optical measurements is normally written as

$$I(\phi, A, \gamma) = A\left[1 + \gamma\,\mathrm{frgn}(\phi)\right], \tag{1}$$

where A is the bias and γ is the contrast. The function $\mathrm{frgn}(\phi)$ is a periodic function, a cosine in the simplest case. In the general case, it contains harmonic terms. The bias and contrast may vary slowly across the field of view: $A = A(x, y)$, $\gamma = \gamma(x, y)$. The bias may also vary with time, if the source intensity is fluctuating: $A = A(t)$. This possibility of bias variation is considered in Sect. 2.2.4. A general example of an intensity pattern is shown in Fig. 1a. In the following, we drop the dependence of I on A and γ, and we simply write the intensity as $I(\phi)$.

Phase-stepping methods require the acquisition of M intensity samples I_k,

$$I_k = I(\phi + k\delta), \tag{2}$$

where $\delta = 2\pi/N$ is the phase shift. In this Section we assume that there is no miscalibration of this phase shift. Phase detection algorithms are most often written as

$$\tilde{\phi} = \arctan\left(\frac{\sum_{k=0}^{M-1} b_k I_k}{\sum_{k=0}^{M-1} a_k I_k}\right), \tag{3}$$

where $\tilde{\phi}$ is the measured phase, lying in the range $[-\pi, \pi]$. Equation (3) can also be interpreted as the computation of the argument of the complex linear combination

$$S(\phi) = \sum_{k=0}^{M-1} c_k I_k, \tag{4}$$

where $c_k = a_k + \mathrm{i}\, b_k$. The intensity can be expanded in a Fourier series as

$$I(\phi) = \sum_{m=-\infty}^{+\infty} \alpha_m \exp(\mathrm{i}\, m\phi), \tag{5}$$

and consequently the sum $S(\phi)$ can also be expanded as

$$S(\phi) = \sum_{m=-\infty}^{+\infty} s_m \exp(\mathrm{i}\, m\phi) = \sum_{m=-\infty}^{+\infty} S_m(\phi). \tag{6}$$

The intensity is real, and so

$$\alpha_m = \alpha_{-m}^* \ . \tag{7}$$

Without loss of generality, one can suppose that $\alpha_1 = \alpha_{-1}$ is real positive. This simply implies that the phase origin is taken at a bright fringe, and the fundamental signal is then proportional to $\cos\phi$.

For designing algorithms, the basic idea is to choose the set of coefficient $\{c_k,\, k = 0 \ldots M-1\}$ so that all components S_m cancel except for S_1. Thus, taking the argument of $S(\phi)$ will directly provide ϕ. Of course, this implies an infinite number of equations, and so requires M to be infinite, which is not possible. However, three considerations have to be taken into account. First, the harmonic $m = -1$ must always be removed, as its amplitude is equal to the $m = 1$ one which is sought. Second, the harmonic $m = 0$, which corresponds to the bias has always a large amplitude (as the intensity is

always positive) and must also be removed. Third, the amplitudes of higher harmonics usually decrease rapidly, and so canceling harmonics of very high order is usually not a critical issue.

So the first choice the user has to make is to choose the extent up to which the harmonics need to be removed. As an example, the user may want to remove harmonics $m = 0$, $m = -1$, $m = \pm 2$ and $m = \pm 3$. The CP method allows an immediate computation of the corresponding set of coefficients $\{c_k\}$.

The CP is related to the coefficients $\{c_k\}$ by

$$P(x) = \sum_{k=0}^{M-1} c_k x^k. \tag{8}$$

From the above equations, it is easy to write the m-th Fourier component of $S(\phi)$:

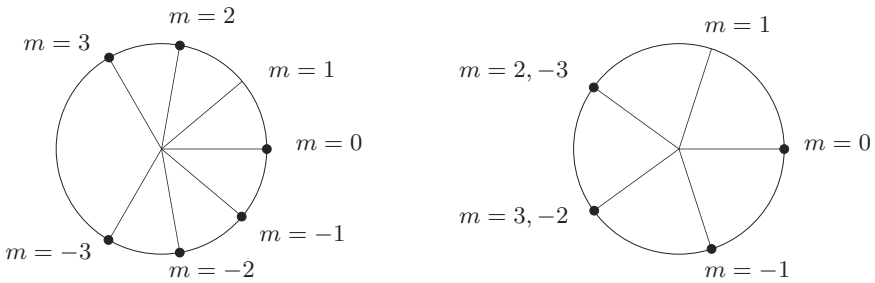
$$\begin{aligned} S_m(\phi) &= \alpha_m \sum_{k=0}^{M-1} c_k \exp[\mathrm{i}m(\phi + k\delta)], \\ &= \alpha_m \exp(\mathrm{i}m\phi) \sum_{k=0}^{M-1} c_k [\exp(\mathrm{i}m\delta)]^k, \\ &= \alpha_m \exp(\mathrm{i}m\phi) P(\zeta^m) \ , \end{aligned} \tag{9}$$

where $\zeta = \exp(\mathrm{i}\delta)$. From this expression, it can be seen that canceling $S_m(\phi)$ requires that $P(\zeta^m) = 0$. This implies that its factorization contains the monomial $x - \zeta^m$. So, if we go back to our example and if the harmonics $m = 0$, $m = -1$ and $m = \pm 2$ are to be removed, the CP should write

$$P(x) = c_{M-1}(x-1)(x-\zeta^{-1})(x-\zeta^2)(x-\zeta^{-2})(x-\zeta^3)(x-\zeta^{-3}), \tag{10}$$

where c_{M-1} can be arbitrarily chosen. Till now, there is no requirement on the value of the phase shift δ. However, a proper choice will minimize the number of necessary samples. To demonstrate this, it is useful to introduce the characteristic diagram, which is the representation in the complex plane of the roots of the CP on the unit circle. If the phase shift is arbitrary, the characteristic diagram corresponding to the polynomial described by 10 will appear as in Fig. 2a. In that case, the polynomial degree is 6, so $M = 7$ intensities are required. If $\delta = 2\pi/5$, the characteristic diagram will appear as in Fig. 2b where there are two roots less. So this choice requires a total of 5 images instead of 7.

If one wants to eliminate the harmonics up to the order j, the general rule is to take $\delta = 2\pi/N$ with $N = j + 2$. The algorithm which is obtained



(a) (b)

Fig. 2. Characteristic diagram corresponding to the polynomial described by (10). (**a**) Arbitrary phase shift (**b**) Optimum phase shift.

in this case is the classical N-bucket one whose CP is

$$\begin{aligned} P_N(x) &= \zeta \prod_{\substack{k=0 \\ k\neq 1}}^{N-1} (x - \zeta^k) \\ &= \zeta \frac{x^N - 1}{x - \zeta} \\ &= 1 + \zeta^{-1}x + \zeta^{-2}x^2 + \ldots + \zeta^{-(N-1)}x^{N-1}. \end{aligned} \tag{11}$$

Indeed, this equation shows that $c_k = \zeta^{-k} = \exp(-\mathrm{i}\, 2k\pi/N)$, that is $a_k = \cos(2k\pi/N)$ and $b_k = -\sin(2k\pi/N)$. Thus the final algorithm can be written as

$$\tilde{\phi} = -\arctan\left[\frac{\sum_{k=0}^{N-1} I_k \sin(2k\pi/N)}{\sum_{k=0}^{N-1} I_k \cos(2k\pi/N)}\right]. \tag{12}$$

Surrel [25] calls this algorithm the DFT one, as it corresponds to the computation of the argument of the first coefficient of the discrete Fourier transform of the set of intensities $\{I_k\}$.

2.2.2 Aliased Algorithms

The constant ζ in the product of monomials in (11) has been introduced in order to obtain the N-bucket algorithm in its classical form. The CP is defined up to an arbitrary constant and changing this constant allows one to

obtain many aliased versions of the same algorithm. From (9), one gets

$$S_1(\phi) = \alpha_1 \exp(\mathrm{i}\phi) P(\zeta). \tag{13}$$

In principle, the algorithm is designed such that $S_1(\phi)$ is the only non negligible component of $S(\phi)$, and taking the argument of $S(\phi)$ yields

$$\tilde{\phi} = \arg[S(\phi)] = \phi + \arg(\alpha_1) + \arg[P(\zeta)]. \tag{14}$$

As mentioned in the preceding section, the argument of α_1 depends on the phase origin: if the phase origin is taken so that a null phase corresponds to a bright fringe, the first Fourier component of the intensity is proportional to $\cos\phi$ and α_1 is real positive (it is real negative if the phase origin is chosen at a dark fringe, as in this case the first Fourier component is proportional to $-\cos\phi$). So, with the phase origin taken at a bright fringe, $\arg(\alpha_1) = 0$ and (14) can be rewritten as

$$\tilde{\phi} = \arg[S(\phi)] = \phi + \arg[P(\zeta)]. \tag{15}$$

It can be seen in this equation that multiplying the CP by an arbitrary coefficient β will result in adding the constant $\arg(\beta)$ to the detected phase. So the phases detected by aliased versions of the same algorithm only differ by a constant.

Aliased algorithms can be found in the literature. As an example, the 3-frame algorithm for $\delta = \pi/2$ [12]

$$\tilde{\phi}_1 = \arctan\left(\frac{I_0 - I_1}{I_1 - I_2}\right), \tag{16}$$

corresponds to the CP $P_1(x) = \mathrm{i} + (1-\mathrm{i})x - x^2 = -(x-1)(x+\mathrm{i})$. One has in this case $P(\zeta) = P(\mathrm{i}) = 2(1+\mathrm{i})$ and following (15), $\tilde{\phi}_1 = \phi + \pi/4$. On the other hand, the algorithm [60]

$$\tilde{\phi}_2 = \arctan\left(\frac{I_0 - 2I_1 + I_2}{I_0 - I_2}\right), \tag{17}$$

corresponds to the CP $P_2(x) = (1+\mathrm{i}) - 2\mathrm{i}x - (1-\mathrm{i})x^2 = (1-\mathrm{i})P_1(x)$. As $P_2(\mathrm{i}) = 4$, the detected phase is $\tilde{\phi}_2 = \phi$. These two algorithms are aliased versions of a single one and correspond to the characteristic diagram presented at line 1 of the table of algorithms in the Appendix. This is not evident from their arctangent form, but is clear from the examination of their respective CPs. The first version given by (16) is more efficient in terms of computing time, as it involves one addition operation less. This is the only reason that can lead one to prefer a particular form. In Sect. 2.2.6 we explain how to choose the multiplying factor of the CP such that a Hermitian symmetry of the coefficients is introduced, allowing us to group the intensity values in pairs, thus reducing the number of multiplication and addition operations.

2.2.3 Self-Calibrating Algorithms

Linear miscalibration. As was mentioned in the introduction, some amount of miscalibration may be present and it is important that the algorithm be designed to be insensitive to this cause of error. A linear miscalibration corresponds to a real phase shift $\delta' = \delta(1+\epsilon) = \delta + \mathrm{d}\delta$. Taking the derivative of $\zeta = \exp(\mathrm{i}\delta)$, one obtains $\mathrm{d}\zeta = \mathrm{i}\zeta\mathrm{d}\delta$ and finally:

$$\begin{aligned}\mathrm{d}P(\zeta^m) &= P'(\zeta^m)m\zeta^{m-1}\mathrm{i}\zeta\mathrm{d}\delta,\\ &= \mathrm{i}m\mathbf{D}P(\zeta^m)\mathrm{d}\delta,\\ &= \mathrm{i}m\delta\epsilon\mathbf{D}P(\zeta^m),\end{aligned} \tag{18}$$

where $P'(x) = \mathrm{d}P(x)/\mathrm{d}x$ and the linear operator $\mathbf{D}$ is defined by

$$\mathbf{D}P(x) = xP'(x). \tag{19}$$

Equation (18) gives the first-order variation of $P(\zeta^m)$. If one wants as before to cancel the term proportional to $\exp(\mathrm{i}m\phi)$ in the sum $S(\phi)$ up to the first order in ϵ, both $P(\zeta^m)$ and $\mathrm{d}P(\zeta^m)$ must vanish. This is equivalent to the condition $P(\zeta^m) = \mathbf{D}P(\zeta^m) = 0$, and it is evident from (19) that $\mathbf{D}P(x) = 0$ for $x \neq 0$ if and only if $P'(x) = 0$.

So, we obtain the major result that the $m-$th harmonic is removed up to the first order in ϵ if ζ^m is a double root of the CP. Clearly, as seen from (18), this condition is not necessary if $m = 0$.

The formula in (18) can be iterated and it is easy to show that an improved insensitivity to miscalibration is obtained for the removal of the harmonic m if ζ^m is a root of increasing order of the CP. More precisely, in the case it is a root of order r, the term in $S(\phi)$ proportional to $\exp(\mathrm{i}m\phi)$ will contain ϵ^r as the term of lowest power in ϵ [25].

Nonlinear miscalibration. We consider in this subsection the case of a nonlinear phase shift

$$\delta_k = k\delta(1 + \epsilon_1 + k\epsilon_2 + k^2\epsilon_3 + \dots). \tag{20}$$

In this case, the m-th Fourier component of the sum $S(\phi)$ can be written as

$$\begin{aligned}S_m(\phi) &= \alpha_m \sum_{k=0}^{M-1} c_k \exp\left[\mathrm{i}m(\phi + \delta_k)\right]\\ &= \alpha_m \exp(\mathrm{i}m\phi) \sum_{k=0}^{M-1} c_k \exp(\mathrm{i}m\delta_k).\end{aligned} \tag{21}$$

In the case of the nonlinear phase shift expressed by (20), the sum in (21) can be written as

$$\begin{aligned}\sum_{k=0}^{M-1} c_k \exp(\mathrm{i}m\delta_k) &= \sum_{k=0}^{M-1} c_k \exp\left[\mathrm{i}mk\delta(1+\epsilon_1+k\epsilon_2+k^2\epsilon_3+\dots)\right] \\ &= \sum_{k=0}^{M-1} c_k[\exp(\mathrm{i}\delta)^m]^k\,[\,1+\mathrm{i}mk\delta\epsilon_1+\mathrm{i}mk^2\delta\epsilon_2+\mathrm{i}mk^3\delta\epsilon_3 \\ &\quad +\dots \\ &\quad +\frac{(\mathrm{i}mk\delta\epsilon_1)^2}{2!}+\frac{(\mathrm{i}mk^2\delta\epsilon_2)^2}{2!}+\frac{(\mathrm{i}mk^3\delta\epsilon_3)^2}{2!}+\dots \\ &\quad +\frac{2(\mathrm{i}mk\delta\epsilon_1)(\mathrm{i}mk^2\delta\epsilon_2)}{2!}+\frac{2(\mathrm{i}mk\delta\epsilon_1)(\mathrm{i}mk^3\delta\epsilon_3)}{2!}+\dots \\ &\quad +\frac{2(\mathrm{i}mk^2\delta\epsilon_2)(\mathrm{i}mk^3\delta\epsilon_3)}{2!}+\dots\,]\,. \end{aligned} \tag{22}$$

It can be seen that the preceding expression involves many different polynomials of $\exp(\mathrm{i}\,m\delta)=\zeta^m$, with coefficients $\{c_k\}$, $\{kc_k\}$, $\{k^2c_k\}$ etc. In terms of a dummy variable x, the first one is the characteristic polynomial $P(x)$. The others are deduced from $P(x)$ by the iterative application of the operator $\mathbf{D}=x\frac{\mathrm{d}}{\mathrm{d}x}$ already introduced in the preceding subsection. So, taking into account (21) and (22), the supplementary term in the sum $S(\phi)$ related to the $m-$th harmonic becomes in the case of a nonlinear phase shift:

$$\begin{aligned}S_m(\phi) = \alpha_m \exp(\mathrm{i}m\phi)\Bigg[&P(\zeta^m) \\ &+\mathrm{i}m\delta\epsilon_1\mathbf{D}P(\zeta^m)+\mathrm{i}m\delta\epsilon_2\mathbf{D}^2P(\zeta^m)+\mathrm{i}m\delta\epsilon_3\mathbf{D}^3P(\zeta^m)+\dots \\ &+\frac{(\mathrm{i}m\delta\epsilon_1)^2}{2!}\mathbf{D}^2P(\zeta^m)+\frac{(\mathrm{i}m\delta\epsilon_2)^2}{2!}\mathbf{D}^4P(\zeta^m)+\frac{(\mathrm{i}m\delta\epsilon_3)^2}{2!}\mathbf{D}^6P(\zeta^m) \\ &+\dots \\ &+\frac{2(\mathrm{i}m\delta\epsilon_1)(\mathrm{i}m\delta\epsilon_2)}{2!}\mathbf{D}^3P(\zeta^m)+\frac{2(\mathrm{i}m\delta\epsilon_1)(\mathrm{i}m\delta\epsilon_3)}{2!}\mathbf{D}^4P(\zeta^m) \\ &+\frac{2(\mathrm{i}m\delta\epsilon_2)(\mathrm{i}m\delta\epsilon_3)}{2!}\mathbf{D}^5P(\zeta^m)+\dots\Bigg]\,. \end{aligned} \tag{23}$$

The conditions driving the insensitivity to a nonlinear phase shift clearly appear in this equation, and consist in cancelling the adequate terms in

$\epsilon_{r_1}^{s_1}\epsilon_{r_2}^{s_2}\ldots\epsilon_{r_t}^{s_t}$, which requires $\mathbf{D}^{r_1 s_1+r_2 s_2+\ldots+r_t s_t}P(\zeta^m)$ to cancel. As mentioned before, $\mathbf{D}^n P(\zeta^m)$ cancels if ζ^m is a root of order $n+1$ of the CP. For example, the quadratic nonlinearity has no incidence up to the first order in ϵ_2 if $\mathbf{D}^2 P(\zeta^m)=0$, that is, if ζ^m is a triple root of the CP. This condition is the same as the one that provides an improved insensitivity to a linear miscalibration, that is, it cancels the term in ϵ_1^2 as well.

2.2.4 Insensitivity to Bias Variation

A bias modulation can be described to a first approximation by a linear variation over the sampling range, as in [30]. So it is assumed that the bias present in intensity I_k includes a term proportional to $k\delta$. This introduces in $S(\phi)$ a supplementary term proportional to $\sum_{k=0}^{M-1} c_k k = P'(1)$. This supplementary term cancels if $P'(1)=0$, that is, if 1 is a double root of the CP [31]. So, this is the necessary and sufficient condition for an algorithm to be insensitive to a linear bias modulation. It is interesting to see how this condition complements what is presented in the preceding section.

If the bias variation is not linear, the order of the root located at $z=1$ must be increased [31].

2.2.5 Windowed-DFT Algorithm

A simple generic algorithm can be derived from the results of the two preceding sections. If one wants to keep the properties of the DFT algorithm in the presence of both miscalibration and bias variation, one has to double all the roots of the DFT algorithm. For the special case of $N=4$, this corresponds to the algorithm presented at line 12 of the table in the Appendix, and the algorithm is written as

$$\phi = \arctan\left(\frac{(I_0-I_6)-3(I_2-I_4)}{2(I_1+I_5)-4I_3}\right). \tag{24}$$

As all the roots are doubled, the CP of the (WDFT) algorithm is proportional to $P_N(x)^2$, where $P_N(x)$ is the CP of the DFT algorithm, as before. In order to obtain the Hermitian symmetry of the coefficients (see Sect. 2.2.6), a multiplying factor ζ^{-1} is introduced so that the polynomial finally becomes

$$P(x) = \zeta^{-1}P_N(x)^2 = \sum_{k=0}^{2N-2} t_k \zeta^{-k-1} x^k, \tag{25}$$

where $\{t_k, k=0,\ldots,2N-2\} = \{1,2,\ldots,N-1,N,N-1,\ldots,2,1\}$. These coefficients can also be interpreted as the triangular windowing of the signal

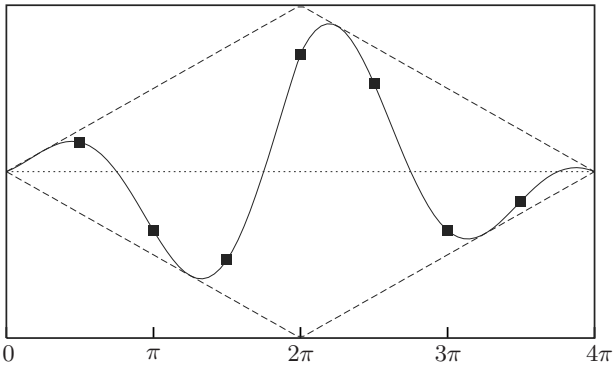


Fig. 3. Triangular windowing performed by the WDFT algorithm

during two periods (Fig. 3). From (25), the WDFT algorithm is written as

$$\phi = \arctan\left[-\frac{\sum_{k=1}^{N-1} k(I_{k-1} - I_{2N-k-1})\sin(2k\pi/N)}{NI_{N-1} + \sum_{k=1}^{N-1} k(I_{k-1} + I_{2N-k-1})\cos(2k\pi/N)}\right]. \tag{26}$$

The total number of samples required by the WDFT algorithm for a phase shift of $2\pi/N$ is $M = 2N - 1$.

It must be recognized that this algorithm will give quite good results in most practical cases, and therefore should be recommended.

2.2.6 Hermitian Symmetry of the Coefficients

The possibility of obtaining aliased algorithms depending on an arbitrary multiplying factor has been presented in Sect. 2.2.2. However, a canonical form can be defined. It is possible to show [32] that the algorithm coefficients can be given the following Hermitian symmetry:

$$c_k^* = c_{M-k-1}, \tag{27}$$

where the star superscript stands for complex conjugation, or equivalently

$$\begin{cases} a_k = a_{M-k-1} \\ b_k = -b_{M-k-1}. \end{cases} \tag{28}$$

This can be achieved when the coefficient c_{M-1} is chosen so that

$$c_0 = c_{M-1}^*, \tag{29}$$

that is, the Hermitian symmetry is obtained for all the coefficients if it is obtained for c_0 and c_{M-1}. Starting from a CP without this symmetry, it is easy to show that it has to be multiplied by

$$\frac{a}{c_0 + c_{M-1}},$$

where a is any real number.

This Hermitian symmetry of the CP coefficients does not confer any specific property to the related algorithm. However, this symmetry is useful for its computer implementation, as it reduces the number of necessary addition and multiplication operations, because it allows us to group intensity terms in pairs, as in the WDFT algorithm presented in (24).

2.2.7 Phasor Representation

It is possible to represent the sum $S(\phi)$ in the complex plane as a sum of vectors, each of which is the representation of a term $c_k I_k$, $k = 0 \ldots M-1$. That graphical representation of a complex number is often called a phasor.

As expressed by (6), the harmonic m builds up the Fourier component $S_m(\phi)$ of $S(\phi)$. As was explained in Sect. 2.2.1, the set of coefficients $\{c_k\}$ should be chosen so that $S_m(\phi) = 0$ for $m \neq 1$. That is, if the signal is $I(\phi) = \exp(i\phi)$, the individual phasors corresponding to all the terms $c_k I_k$ add constructively so that the resulting phasor has a nonvanishing length. On the other hand, other harmonics should build up the null vector. In this case, the phasor representation of the sum $S_m(\phi)$ is a closed path, starting and ending at the origin.

We present here three examples. The first one corresponds to the DFT (or N-bucket) algorithm with $\delta = \pi/3$. The characteristic diagram for this algorithm is displayed at line 10 of the table in the Appendix. From the examination of the location of the roots , one can deduce that this algorithm is designed to cancel harmonics $m = 0, -1, \pm 2, \pm 3, \pm 4$. The phasor construction of the Fourier components $S_m(\phi)$ is shown in Fig. 4.

The second example corresponds to the WDFT algorithm with $\delta = \pi/3$, that is $N = 6$ and $M = 11$. The characteristic diagram for this algorithm is displayed at line 23 of the table in the Appendix. As in the example above, this algorithm is designed to cancel harmonics $m = 0, -1, \pm 2, \pm 3, \pm 4$, but this time with insensitivity to miscalibration. The phasor construction of the Fourier components $S_m(\phi)$ is shown in Fig. 5.

The last example corresponds to the 9-sample algorithm [33]

$$\phi = \arctan\left[\frac{I_0 - I_8 - 2(I_1 - I_7) - 14(I_2 - I_6) - 18(I_3 - I_5)}{2(I_0 + I_8) + 8(I_1 + I_7) + 8(I_2 + I_6) - 8(I_3 + I_5) - 20I_4}\right], \quad (30)$$

whose diagram is presented at line 21 of the table in the Appendix. The phasor construction of the Fourier components $S_m(\phi)$ is shown in Fig. 6. It

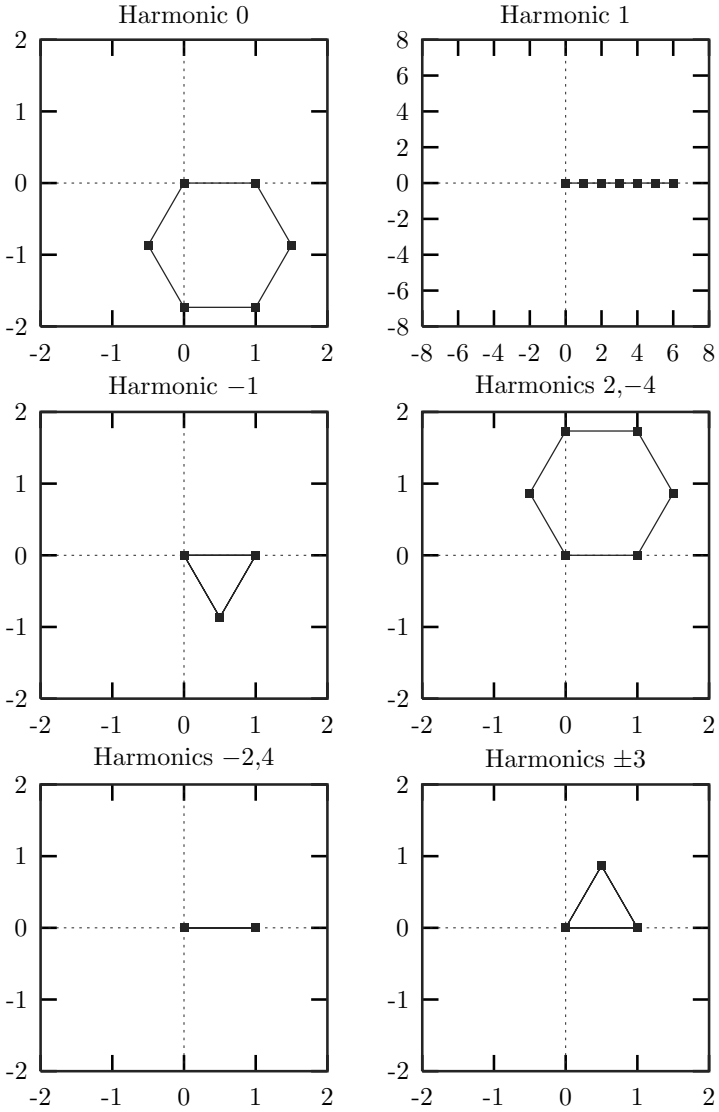


Fig. 4. Phasor representation of the Fourier components of $S(\phi)$ for the DFT algorithm with $\delta = \pi/3$. The characteristic diagram of this algorithm is displayed at line 10 of the table in the Appendix

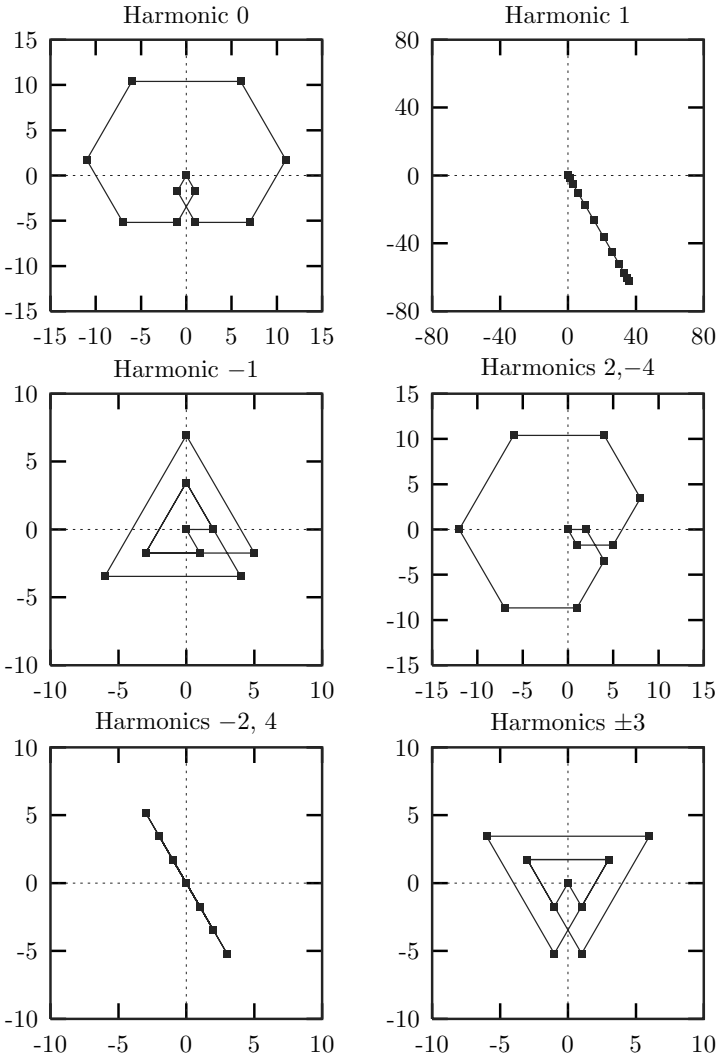


Fig. 5. Phasor representation of the Fourier components of $S(\phi)$ for the windowed-DFT algorithm with $\delta = \pi/3$. The characteristic diagram of this algorithm is displayed at line 23 of the table in the Appendix

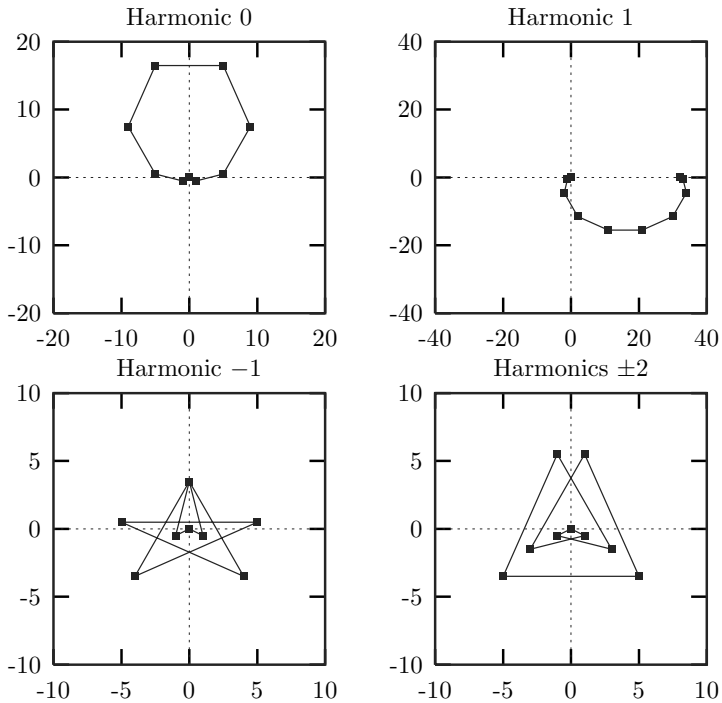


Fig. 6. Phasor representation of the Fourier components of $S(\phi)$ for the 9-sample algorithm described by (30). The characteristic diagram of this algorithm is displayed at line 21 of the table in the Appendix

can be seen that for this algorithm, the construction of $S_1(\phi)$ is not a straight line. That means that the recorded samples are not used in an efficient way. The consequence is that the SNR or "signal strength" will be poor. This effect will be detailed in Sect. 2.5.2.

2.2.8 Relation with the Fourier Transform Approach

The correspondence between our approach and the so-called Fourier transform (FT) approach [24,34] is straightforward. In the FT approach, the sampling functions related to the numerator and denominator of the fraction defining the tangent of the measured phase are introduced:

$$\begin{cases} f_1(\phi) = \sum_{k=0}^{M-1} b_k \delta(\phi + k\delta) \\ f_2(\phi) = \sum_{k=0}^{M-1} a_k \delta(\phi + k\delta) \ . \end{cases} \tag{31}$$

The plus sign is chosen in front of the phase shifts $k\delta$ to be consistent with (2). Actually, in the original work [24], those sampling functions are written as function of time, as it is supposed that the phase shift varies linearly with time. The correspondence is given by

$$\phi = 2\pi\nu_s t, \tag{32}$$

where ν_s is the signal frequency. Then we consider the Fourier transforms $\mathcal{F}$ of the sampling functions f_1 and f_2 and we use ν as the conjugated Fourier variable:

$$\begin{cases} F_1(\nu) = (\mathcal{F} f_1)(\nu) = \sum\limits_{k=0}^{M-1} b_k \exp(\mathrm{i}\, 2k\pi\nu\delta) \\ F_2(\nu) = (\mathcal{F} f_2)(\nu) = \sum\limits_{k=0}^{M-1} a_k \exp(\mathrm{i}\, 2k\pi\nu\delta). \end{cases} \tag{33}$$

The coefficients $\{a_k\}$ and $\{b_k\}$ being real, it can be seen that:

$$\begin{cases} F_1(-\nu) \ = F_1^*(\nu) \\ F_2(-\nu) = F_2^*(\nu) \ . \end{cases} \tag{34}$$

Still denoting $\zeta = \exp(\mathrm{i}\,\delta)$, we obtain

$$\begin{aligned} F_2(\nu) + \mathrm{i}\, F_1(\nu) &= \sum_{k=0}^{M-1} c_k \exp(\mathrm{i}\, 2k\pi\nu\delta), \\ &= P\left(\zeta^m\right), \end{aligned} \tag{35}$$

where $P(x)$ is the CP of the considered algorithm and $m = 2\pi\nu$ is the angular frequency of ϕ. Taking (34) into account, one can write

$$P(\zeta^{-m}) = F_2(-\nu) + \mathrm{i}\, F_1(-\nu) = F_2^*(\nu) + \mathrm{i}\, F_1^*(\nu). \tag{36}$$

Finally, we obtain

$$\begin{aligned} P\left(\zeta^m\right) &= F_2(\nu) + \mathrm{i}\, F_1(\nu) \\ P^*\left(\zeta^{-m}\right) &= F_2(\nu) - \mathrm{i}\, F_1(\nu). \end{aligned} \tag{37}$$

From this , all the results from the CP theory can be converted in terms of the FT theory, and vice versa. In particular, the elimination of harmonics m and $-m$ corresponds to $P(\zeta^m) = P(\zeta^{-m}) = P^*(\zeta^{-m}) = 0$, that is

$$\begin{aligned} F_2(\nu) + \mathrm{i}\, F_1(\nu) &= 0, \\ F_2(\nu) - \mathrm{i}\, F_1(\nu) &= 0. \end{aligned} \tag{38}$$

This is equivalent to

$$F_1(\nu) = F_2(\nu) = 0, \tag{39}$$

which is the condition given by the FT theory. The detection of the harmonic $+m$ requires us to eliminate the $-m$ component only, that is to cancel $P(\zeta^{-m})$, and so the condition is written as

$$F_1(\nu) = -\mathrm{i}\, F_2(\nu), \tag{40}$$

which is the form given by the FT theory, except for the sign. This discrepancy comes from the sign chosen in front of the phase shifts.

2.2.9 Relation with Averaging and Data Windowing

The averaging technique. The averaging technique has been first introduced by *Schwider* et al. [20] and was systematically extended and investigated by *Schmit* and *Creath* [26,27]. The data windowing technique was proposed by *de Groot* [28] (and developed to its ultimate extent by the same author [35]!). In this section we show that the extended averaging and the data windowing techniques can be interpreted using the CP as well.

Schwider et al. [20] noticed that the phase errors in phase-shifting measurements are very often modulated at twice the frequency of the phase ϕ. That is, the measured phase $\tilde{\phi}$ can be written as $\tilde{\phi} = \phi + a\,\sin(2\phi + \psi)$. So, they proposed to perform two sets of measurements, shifted by a quarter of a period, in order to get two phases $\tilde{\phi}_1 = \phi + a\,\sin(2\phi + \psi)$ and $\tilde{\phi}_2 = \phi + \pi/2 + a\,\sin(2\phi + \pi + \psi)$, so that $(\tilde{\phi}_1 + \tilde{\phi}_2 - \pi/2)/2 = \phi$, thus eliminating the phase error. They wrote the arctangent form of an algorithm

$$\tilde{\phi} = \arctan\left(\frac{N}{D}\right). \tag{41}$$

and they showed that, instead of computing two phases $\tilde{\phi}_1$ and $\tilde{\phi}_2$ from (41) and averaging them, it was more efficient to use

$$\tilde{\phi} = \arctan\left(\frac{N_1 + N_2}{D_1 + D_2}\right). \tag{42}$$

This can be applied to any M-sample algorithm but, if the phase shift of the algorithm is $\pi/2$, the two sample sets overlap each other so that the resulting algorithm requires only $M + 1$ samples.

Schmit and *Creath* [26] have extended this idea and proposed to iterate the procedure, that is, to average two $(M + 1)$-sample algorithms obtained as before to get an $(M + 2)$-sample algorithm:

$$\begin{aligned}\tilde{\phi} &= \arctan\left[\frac{(N_1 + N_2) + (N_2 + N_3)}{(D_1 + D_2) + (D_2 + D_3)}\right], \\ &= \arctan\left(\frac{N_1 + 2N_2 + N_3}{D_1 + 2D_2 + D_3}\right).\end{aligned} \tag{43}$$

In a recent paper, the same authors propose the multiple averaging technique [27], in which not only two but three or more algorithms are averaged. In the latter paper, a large number of new algorithms are derived and analysed in terms of data windowing. As examples, the authors investigated different 8-sample algorithms obtained by averaging many different algorithms in different ways. They showed that the result of this averaging process is the windowing of the intensity data set. However, very similar windowing functions turn out to correspond to algorithms that have very different properties.

In a paper by *de Groot* [28], the reverse approach is used, that is, a data window function is given and then the algorithm is derived. Here too, only an *a posteriori* study is capable of elucidating the properties of the obtained algorithm.

Algebraic interpretation of a shift of the data set. Here, we investigate how a shift of the data set (i.e. using the set $\{I_n, I_{n+1} \dots I_{n+M-1}\}$ instead of $\{I_0, I_1 \dots I_{M-1}\}$ as the algorithm input) can be taken into account using the CPs. We will concentrate on algorithms which use a $\pi/2$ phase shift. In any case, results can be straightforwardly extended to the case of a generic phase shift δ.

The N/D notation in (41) introduced by *Schwider* et al. [20] can evidently be related to the sum $S(\phi)$ by (we drop the dependence on ϕ)

$$S = D + i\,N. \tag{44}$$

When two sets of samples are used as in (42), the sum S is written as

$$S = S_1 + S_2. \tag{45}$$

The first term of the sum has the argument ϕ and the second one the argument $\phi + \phi/2$, because the first sample it involves has already a $\pi/2$ phase lag. Both terms have the same magnitude, so that the argument of the sum S is $\phi + \pi/4$. The complex sum corresponding to the second set of samples is

$$S_2 = \sum_{k=1}^{M} c_k I_k \ . \tag{46}$$

Hence the corresponding CP is

$$\sum_{k=1}^{M} c_k x^k = x\,P_1(x) \ , \tag{47}$$

where $P_1(x)$ is the CP corresponding to the processing of the first data set. In the same way, using the algorithm formula with a third set of samples starting at intensity I_2 to construct a sum S_3 would correspond to the polynomial $x^2\,P_1(x)$, and so on.

If the sum S_{n+1}, which has the argument $n\pi/2$, is multiplied by $(-\mathrm{i})^n$, one obtains a complex number of argument ϕ, and a sum of such terms $(-\mathrm{i})^n S_{n+1}$ will also have the argument ϕ. The characteristic polynomial corresponding to $(-\mathrm{i})^n S_{n+1}$ is

$$P_{n+1}(x) = (-\mathrm{i}\,x)^n P_1(x). \tag{48}$$

This result is the algebraic representation of a shift of the input data set.

Example. As an example, let us take an algorithm with $\delta = \pi/2$, whose CP is denoted by $P_1(x)$. Following (48), when two sample sets are averaged the corresponding polynomial will be

$$\begin{aligned} Q_1(x) &= P_1(x) + P_2(x), \\ &= P_1(x)(1 - \mathrm{i}\,x), \\ &= -\mathrm{i}\,P_1(x)(x + \mathrm{i}). \end{aligned} \tag{49}$$

The improvement is evident, as $-\mathrm{i}$ is now a root of higher order, thus increasing the algorithm insensitivity to miscalibration. Now, if three sample sets are averaged, one obtains

$$\begin{aligned} R(x) &= P_1(x) + P_2(x) + P_3(x), \\ &= P_1(x)(1 - \mathrm{i}\,x - x^2), \\ &= -P_1(x)\frac{x^3 + \mathrm{i}}{x - \mathrm{i}}. \end{aligned} \tag{50}$$

The polynomial expressed by the last fraction has as roots the 3rd roots of $-\mathrm{i}$, except for i, that is $\exp(-\mathrm{i}\,\pi/6)$ and $\exp(-\mathrm{i}\,5\pi/6)$. Here, no improvement is obtained with respect either to miscalibration or to harmonics insensitivity. Thus the averaging process cannot ascertain by itself that an interesting algorithm will be obtained. Similarly, the windowing technique (that is, choosing a windowing function using usual ideas of signal processing) will not always provide an efficient algorithm. Moreover, the properties of the obtained algorithms will have to be found by a thourough a posteriori study. So, there is no real advantage of these approaches over the CP method.

A detailed investigation of the various algorithms proposed by *Schmit* and *Creath* [27] is presented in [36].

2.2.10 Relation with the Recursion Method

The recursion method has been introduced recently by *Phillion* [29]. In Sect. 4.F of [29], the author shows some equivalence of his method with the CP one. We extend and complement here his results.

First, starting from the coefficients $\{c_k\}$ of the algorithm (denoted $\{r_j\}$ in the original paper), the weights $\{w_k\}$ are defined by

$$w_k = c_k \exp(\mathrm{i}\,k\delta) = c_k \zeta^k. \tag{51}$$

In Sect. 4.F of [29], the author introduces the $w-$polynomial $p(x)$ by

$$p(x) = \sum_{k=0}^{M-1} w_k x^k. \tag{52}$$

It is clear from (51) and (52) that

$$p(x) = P(\zeta x), \tag{53}$$

where $P(x)$ is still the algorithm's CP. All results in Phillion's paper [29] can be analysed using this relation.

For example, the author introduces recursion relations by

$$\overline{w}_k = \sum_{l=0}^{R} a_k w_{l+k}, \tag{54}$$

and he notices that this is equivalent to multiplying the CP by

$$Q(x) = \sum_{l=0}^{R} a_l (\zeta^{-1} x)^l. \tag{55}$$

The recursion relations that are presented correspond to introducing more roots in the CP. For example, the recursion relation for $\delta = \pi/2$, $\overline{w}_k = w_k + w_{k+1}$, is equivalent to multiplying the CP by $Q(x) = 1 - \mathrm{i}\, x = -\mathrm{i}(x + \mathrm{i})$. This increases by one the order of multiplicity of the root located at $-\mathrm{i}$.

It is interesting to compare this with the averaging technique, especially with (49). In general, the recursion and averaging techniques seem to be very closely related.

Also, a "figure of merit" N_{F} is introduced by Phillion

$$N_{\mathrm{F}} = \frac{\left| \sum_k w_k \right|}{\left(\sum_k |w_k|^2 \right)^{1/2}}. \tag{56}$$

Owing to (51), the sum in the numerator is nothing else than $P(\zeta)$, and so it can be seen that

$$N_{\mathrm{F}} = \eta \sqrt{M}, \tag{57}$$

where η is the efficiency factor defined by (87) in Sect. 2.5.2. However, it should be noticed that $0 \leq \eta \leq 1$ for any algorithm, and therefore the factor η allows more straightforward comparisons. Other results in Phillion's paper concerning the noise evaluation are the same as those presented in Sect. 2.5.

2.3 Global Phase Detection

The global phase detection technique is called the Fourier transform technique [1,2,3,4,5,6,7,8], but it should not be confused with the technique of the same name presented in Sect. 2.2.8. It is called "global" as the whole set of intensity data in the fringe pattern is used to calculate the whole set of phase data.

The requirement on the fringe pattern is that the useful information should appear as the modulation of a spatial carrier of frequency $\boldsymbol{f}_0$, i.e. the intensity signal should appear as

$$\begin{aligned} I(\boldsymbol{r}) &= A\left\{1+\gamma\cos\left[2\pi\boldsymbol{f}_0\boldsymbol{r}+\phi(\boldsymbol{r})\right]\right\}, \\ &= A+\frac{A\gamma}{2}\exp(\mathrm{i}\,2\pi\boldsymbol{f}_0\boldsymbol{r})\exp[\mathrm{i}\,\phi(\boldsymbol{r})] \\ &\quad +\frac{A\gamma}{2}\exp(-\mathrm{i}\,2\pi\boldsymbol{f}_0\boldsymbol{r})\exp[-\mathrm{i}\,\phi(\boldsymbol{r})], \\ &= A+C(\boldsymbol{r})\exp(\mathrm{i}\,2\pi\boldsymbol{f}_0\boldsymbol{r})+C^*(\boldsymbol{r})\exp(-\mathrm{i}\,2\pi\boldsymbol{f}_0\boldsymbol{r}), \end{aligned} \tag{58}$$

where

$$C(\boldsymbol{r}) = \frac{A\gamma}{2}\exp[\mathrm{i}\,\phi(\boldsymbol{r})]. \tag{59}$$

Here we suppose that the fringe intensity profile is sinusoidal. We will mention briefly below how things are modified if the signal contains harmonics.

Taking the Fourier transform of (58), we can write

$$\begin{aligned} \widehat{I}(\boldsymbol{f}) &= A\delta(\boldsymbol{f})+\widehat{C}(\boldsymbol{f}-\boldsymbol{f}_0)+\widehat{C^*}(\boldsymbol{f}+\boldsymbol{f}_0), \\ &= A\delta(\boldsymbol{f})+\widehat{C}(\boldsymbol{f}-\boldsymbol{f}_0)+\widehat{C}^*(-\boldsymbol{f}-\boldsymbol{f}_0). \end{aligned} \tag{60}$$

The function $C(\boldsymbol{r})$ is now supposed to have only a low frequency content, i.e. the useful signal $\phi(\boldsymbol{r})$ is supposed to vary slowly with respect to the carrier frequency $\boldsymbol{f}_0$. This means that $\widehat{C}(\boldsymbol{f})$ is concentrated around the origin. This is shown in Fig. 7 in the special case of $\boldsymbol{f}_0 = f_0\boldsymbol{x}$, where $\boldsymbol{x}$ is the unit vector along direction x. Thus from (60) it can be seen that $\widehat{I}(\boldsymbol{f})$ will consist of three distinct parts: a Dirac delta function corresponding to the mean value of $I(\boldsymbol{r})$ and two side-lobes corresponding to the spectrum of $C(\boldsymbol{r})$ translated by the amount of the carrier frequency, $\boldsymbol{f}_0$ and $-\boldsymbol{f}_0$ (Fig. 8).

Therefore it is possible to band pass filter the intensity spectrum so that only the right-hand side lobe $\widehat{C}(\boldsymbol{f}-\boldsymbol{f}_0)$ is retained (Fig. 9). The resulting function is approximated by

$$g(\boldsymbol{f}) = \widehat{C}(\boldsymbol{f}-\boldsymbol{f}_0)\ . \tag{61}$$

The inverse Fourier transform of this function is $C(\boldsymbol{r})\exp(\mathrm{i}\,2\pi\boldsymbol{f}_0\boldsymbol{r})$. Following (59), the last step is then to take the argument of that result to obtain

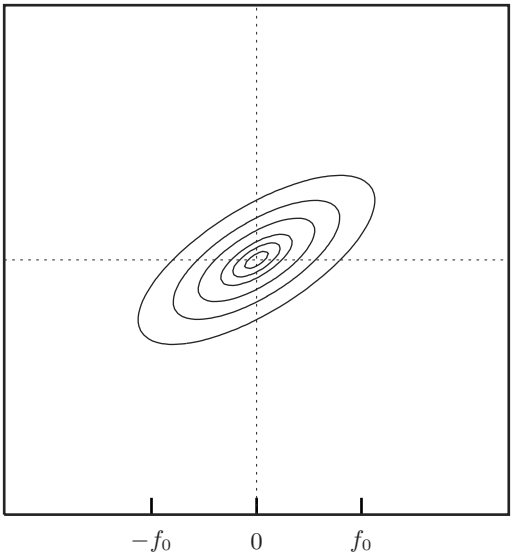


Fig. 7. Low frequency spectrum $|\widehat{C}(\boldsymbol{f})|$ containing the signal information

$\phi(\boldsymbol{r}) + 2\pi \boldsymbol{f}_0 \boldsymbol{r}$. Then the linear carrier phase variation can be easily subtracted. An equivalent method is to translate the side lobe to the origin, to obtain $\widehat{C}(\boldsymbol{f})$ before taking its inverse Fourier transform. Practically, the filtering process in the Fourier space should be smoother than a simple stepwise windowing, because it is well known that such a filtering introduces in the original space parasitic ripples that will distort the phase.

It will often be very difficult to separate perfectly the useful signal frequencies. First, the signal profile may not be perfectly sinusoidal and so other side lobes may be present at frequencies $\pm n\boldsymbol{f}_0$, where n is integer. Thus some crosstalk between lobes may be present, mixing fundamental and harmonic information.

The Fourier transform method can be used when two orthogonal carriers encode two different sets of information [37]. A possible example is when a cross-grid is used for measuring displacements [38]. Each grid-line direction is a carrier used for the detection of the displacement component orthogonal to that direction. The carrier frequency vectors can be written $\boldsymbol{f}_0$ and $\boldsymbol{f}_1$. In the case of a nonsinusoidal signal, the spectrum will appear as in Fig. 10. The required passing band for the detection of the information encoded by the carrier $\boldsymbol{f}_0$ appears in Fig. 11.

However, it can be understood that the filtering process cannot be easily automated, as the filtered domain will depend on the shape of the spectrum. The bias itself may vary across the field, thus widening the central Dirac function with also some risk of cross talk with the useful frequencies. The

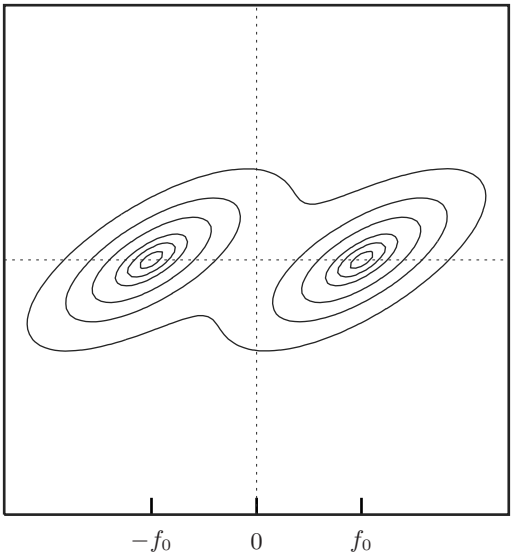


Fig. 8. Spectrum of the intensity, with its two side-lobes centered at $\boldsymbol{f}_0$ and $-\boldsymbol{f}_0$ (the Dirac function at point (0,0) is not represented)

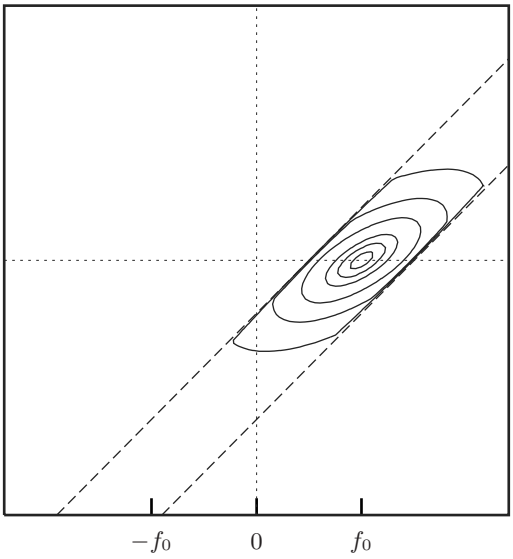


Fig. 9. Bandpass filtered spectrum of the intensity retaining the side-lobe centered at $\boldsymbol{f}_0$

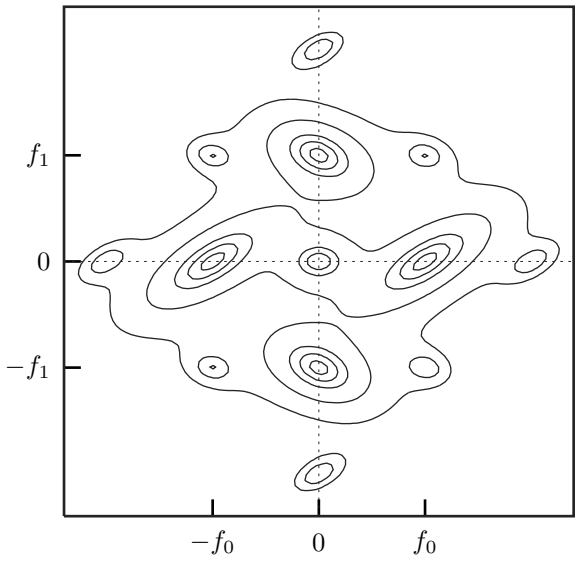


Fig. 10. Intensity spectrum in the case of a non-sinusoidal signal with two orthogonal carriers

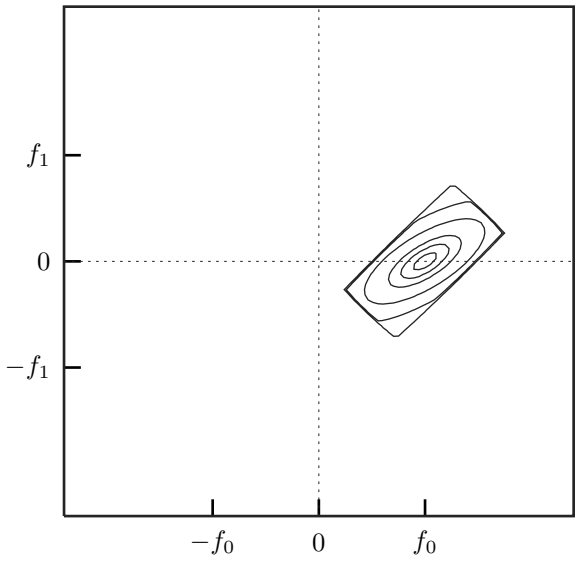


Fig. 11. New definition of the bandpass filter for the spectrum appearing in Fig. 10

fundamental problem is that the resulting errors on the phase are almost impossible to quantify in the general case.

The major advantage of this technique over spatial phase-stepping is that the available range of the fringe frequency is wider. If we consider $[f_0/2, 3f_0/2]$ as a typical band-pass filter width, the local fringe frequency can lie within that range. So the local 'miscalibration', that is the local mismatch between the actual and the carrier frequencies, can be within the range $[-50\%, 33\%]$. However, this range is very theoretical, and in practice the requirements on the fringe frequency variations across the field may not be so different between the Fourier transform and the spatial phase-stepping methods. However, this issue has not yet been precisely addressed.

It is sometimes thought that it is not necessary that the carrier frequency be high. This overlooks a major feature of the phase modulation technique. Indeed, using a low-frequency carrier will decrease the spatial resolution, because it can be considered that the "gauge length" is at least equal to one signal period and so the effective spatial resolution is the carrier period and not the pixel. So even with the Fourier transform method it should be considered desirable that the largest number of fringes appear in the field.

2.4 Residual Phase Errors

A crucial point for real engineering applications is the possibility of evaluating the sensitivity and precision. One requirement is that the phase errors can be evaluated. The CP theory gives analytic formulae for almost any source of systematic errors.

2.4.1 General Evaluation

The residual errors in the phase measurement can easily be evaluated using the CP theory. Normally, the sum $S(\phi)$ should contain $S_1(\phi)$ as the predominant term. From (23) we have

$$S_1 = S_{10} + \Delta S_1 \ , \tag{62}$$

where

$$S_{10} = \alpha_1 \exp(\mathrm{i}\,\phi) P(\zeta), \tag{63}$$

and where ΔS_1 is a small quantity involving possible miscalibration terms. In the ideal case $\Delta S_1 = 0$ and the phase detected is $\phi + \psi$, where $\psi = \arg[P(\zeta)]$. This was already stated by (15) in Sect. 2.2.2. When harmonics or

miscalibration are present, the sum S can be written

$$\begin{aligned} S &= S_{10} + \Delta\, S\ , \\ &= S_{10}\left(1 + \frac{\Delta\, S}{S_{10}}\right)\ , \\ &= S_{10}\left[1 + \Re\left(\frac{\Delta\, S}{S_{10}}\right) + i\,\Im\left(\frac{\Delta\, S}{S_{10}}\right)\right]\ , \\ &\approx |S_{10}|\exp[\mathrm{i}(\phi+\psi)]\left[1 + \mathrm{i}\,\Im\left(\frac{\Delta\, S}{S_{10}}\right)\right]\ , \end{aligned} \tag{64}$$

where $\Re$ and $\Im$ are the real-part and imaginary-part operators. On the other hand,

$$S = |S|\exp[\mathrm{i}\,(\phi+\psi+\Delta\,\phi)] \approx |S|\exp[\mathrm{i}(\phi+\psi)](1+\mathrm{i}\,\Delta\,\phi). \tag{65}$$

So, using $|S| = |S_{10}| + \Delta\,|S|$, the phase error can be written

$$\Delta\,\phi = \Im\left(\frac{\Delta\, S}{S_{10}}\right). \tag{66}$$

We see from the above equation and from (63) that the phase error formula will always involve a division by $\alpha_1 P(\zeta)$. So, it is a good idea to introduce the *normalized* CP as

$$\tilde{P}(x) = \frac{P(x)}{P(\zeta)} \tag{67}$$

and the normalized amplitude of the m-th harmonic $\tilde{\alpha}_m = \alpha_m/\alpha_1$. Now, it is very easy to give an analytic formula for the phase error when miscalibration and/or harmonics are present. It can be directly deduced from (23) and (66) and yields

$$\begin{aligned} \Delta\,\phi = \sum_m \Im\Bigg\{\tilde{\alpha}_m \exp[\mathrm{i}(m-1)\phi]\Bigg[\tilde{P}(\zeta^m) + \\ + \mathrm{i}m\delta\epsilon_1\mathbf{D}\tilde{P}(\zeta^m) + \mathrm{i}m\delta\epsilon_2\mathbf{D}^2\tilde{P}(\zeta^m) + \mathrm{i}m\delta\epsilon_3\mathbf{D}^3\tilde{P}(\zeta^m)\ldots \\ + \frac{(\mathrm{i}m\delta\epsilon_1)^2}{2!}\mathbf{D}^2\tilde{P}(\zeta^m) + \frac{(\mathrm{i}m\delta\epsilon_2)^2}{2!}\mathbf{D}^4\tilde{P}(\zeta^m) + \frac{(\mathrm{i}m\delta\epsilon_3)^2}{2!}\mathbf{D}^6\tilde{P}(\zeta^m) \\ + \ldots + \frac{2(\mathrm{i}m\delta\epsilon_1)(\mathrm{i}m\delta\epsilon_2)}{2!}\mathbf{D}^3\tilde{P}(\zeta^m) + \frac{2(\mathrm{i}m\delta\epsilon_1)(\mathrm{i}m\delta\epsilon_3)}{2!}\mathbf{D}^4\tilde{P}(\zeta^m) \\ + \frac{2(\mathrm{i}m\delta\epsilon_2)(\mathrm{i}m\delta\epsilon_3)}{2!}\mathbf{D}^5\tilde{P}(\zeta^m) + \ldots\Bigg]\Bigg\}. \end{aligned} \tag{68}$$

Notice that in this sum the term $\tilde{\alpha}_1 \tilde{P}(\zeta)$ appears for $m = 1$. Obviously, it is equal to 1 and vanishes when taking the imaginary part, as it should.

Two examples illustrating the application of (68) are presented below.

2.4.2 Effect of the m-th Harmonic

As a first example, let us consider a nonsinusoidal signal and the effects of the m-th harmonic. As stated by (7) of Sect. 2.2.1, both harmonics m and $-m$ are present with conjugated amplitudes, due to the fact that the intensity is a real signal. From (9), we see that the corresponding terms in $S(\phi)$ are $\alpha_m \exp(\mathrm{i}\, m\phi) P(\zeta^m) + \alpha_m^* \exp(-\mathrm{i}\, m\phi) P(\zeta^{-m})$. From (68), the phase error that results is

$$\Delta\, \phi = \Im\{\tilde{\alpha}_m \exp[\mathrm{i}(m-1)\phi]\tilde{P}(\zeta^m) + \tilde{\alpha}_m^* \exp[-\mathrm{i}(m+1)\phi]\tilde{P}(\zeta^{-m})\}. \quad (69)$$

This error has two components at $(m-1)$ and $(m+1)$ times the fringe frequency.

2.4.3 Effect of the Miscalibration

As a second example, we consider the presence of linear miscalibration, and a purely sinusoidal signal. In that case, only $m = 1$ and $m = -1$ must be considered, with $\epsilon_1 = \epsilon \neq 0$. We take as an example an algorithm that is not self-calibrating. From the results of Sect. 2.2.3, that means $\mathbf{D}P(\zeta^{-1}) \neq 0$. As was already mentioned after (7) in Sect. 2.2.1, $\alpha_1 = \alpha_{-1}$. Taking into account $P(\zeta^{-1}) = 0$, (68) gives for the phase error

$$\Delta\, \phi = \delta\epsilon\ \Im[\mathrm{i}\, \mathbf{D}\tilde{P}(\zeta) - \mathrm{i}\ \exp(-\mathrm{i}\, 2\phi)\mathbf{D}\tilde{P}(\zeta^{-1})]\ . \quad (70)$$

So there is a static error and a modulation at twice the fringe frequency. Let us consider as an example the DFT algorithm. In this case, one has [25]

$$\begin{cases} \mathbf{D}\tilde{P}(\zeta) = \dfrac{N-1}{2} \\ \mathbf{D}\tilde{P}(\zeta^{-1}) = \dfrac{\mathrm{i}\ \exp(\mathrm{i}\,\delta)}{2\sin\delta}. \end{cases}$$

Thus the phase error can be written for this algorithm as

$$\Delta\, \phi = \delta\epsilon \left[\frac{N-1}{2} - \frac{\sin(2\phi - \delta)}{2\sin\delta}\right]\ , \quad (71)$$

which corresponds to the formula found in the literature [39,25].

In the case of a self-calibrating algorithm with, say, $\mathbf{D}P(\zeta^{-1}) = 0$ and $\mathbf{D}^2 P(\zeta^{-1}) \neq 0$, (70) would be changed into

$$\Delta\, \phi = \delta\epsilon\ \Im\left[\mathrm{i}\, \mathbf{D}\tilde{P}(\zeta) - \frac{\delta\epsilon}{2}\ \exp(-\mathrm{i}\, 2\phi)\mathbf{D}^2\tilde{P}(\zeta^{-1})\right]. \quad (72)$$

It should be clear now how all possible cases can be handled using (68). Other examples of phase error evaluations can be found in [32].

2.5 Effect of Additive Noise

For metrological purposes, it is very important that the sensitivity can be evaluated. A good measurement of the sensitivity is from the standard deviation of the phase noise obtained after detection. A general formula can be given which shows the respective influence of the different parameters at the acquisition and processing levels. Further details of the derivation of the results presented in this section can be found in [40].

2.5.1 Average of the Phase Noise

In this section, the intensity signal is supposed to be a perfect sine:

$$I = A(1 + \gamma \cos\phi).$$

In this case the Fourier coefficient α_1 is

$$\alpha_1 = \frac{A\gamma}{2}. \tag{73}$$

Also, the sampling spacing is assumed to be perfect, that is $\epsilon = 0$, so that the sum $S(\phi)$ can be written as

$$S(\phi) = S_1(\phi) = S_{10}(\phi) = |S_{10}| \exp(\mathrm{i}\phi), \tag{74}$$

where $S_{10}(\phi)$ has the same meaning as in Sect. 2.4.1. From (9) and (73) we get

$$|S_1| = \frac{A\gamma}{2}|P(\zeta)|. \tag{75}$$

We now examine the effect on the phase measurement of additive noise present in the intensity signal [40]. Each recorded intensity I_k is supposed to contain an additional centered noise ΔI_k of variance σ^2, supposed not to depend on k:

$$I_k = A\left[1 + \gamma \cos\left(\phi + k\delta\right)\right] + \Delta I_k. \tag{76}$$

It is also assumed that there is a statistical independence of the noise corresponding to two different values of intensity, i.e.

$$\langle \Delta I_k \Delta I_m \rangle = \sigma^2 \delta_{km}, \tag{77}$$

where δ_{km} is the Kronecker delta. The signal-to-noise ratio of the signal is

$$\mathrm{SNR} = \frac{A\gamma}{\sigma\sqrt{2}}\ . \tag{78}$$

In the same way as in Sect. 2.4.1, the phase perturbation can be written as

$$\Delta\,\phi = \Im\left(\frac{\Delta\,S}{S_{10}}\right). \tag{79}$$

Let us define two quantities J and K by

$$\frac{\Delta\,S}{|S_{10}|} = J + \mathrm{i}\,K, \tag{80}$$

so that

$$J = \sum_{k=0}^{M-1} a_k \frac{\Delta\,I_k}{|S_{10}|}\ , \tag{81}$$

$$K = \sum_{k=0}^{M-1} b_k \frac{\Delta\,I_k}{|S_{10}|}\ . \tag{82}$$

Then it is easy to evaluate $\Delta\,\phi$ from (79) and (80) as

$$\Delta\,\phi = K\cos\phi - J\sin\phi. \tag{83}$$

Hence as $\langle J\rangle = \langle K\rangle = 0$, the average phase error cancels, i.e.

$$\langle\Delta\,\phi\rangle = 0\ . \tag{84}$$

2.5.2 Variance of the Phase Noise

The variance of the phase error can be evaluated from (83) squared and averaged:

$$\begin{aligned}\left\langle\Delta\,\phi^2\right\rangle &= \left\langle K^2\right\rangle\cos^2\phi + \left\langle J^2\right\rangle\sin^2\phi - 2\left\langle JK\right\rangle\sin\phi\cos\phi\ ,\\ &= \frac{1}{2}\left[\left\langle J^2\right\rangle + \left\langle K^2\right\rangle - \left(\left\langle J^2\right\rangle - \left\langle K^2\right\rangle\right)\cos(2\phi) - 2\left\langle JK\right\rangle\sin(2\phi)\right].\end{aligned} \tag{85}$$

From the algorithm coefficients, three parameters r, β and θ can be defined by

$$\sum_{k=0}^{M-1}|c_k|^2 = \sum_{k=0}^{M-1}(a_k^2 + b_k^2) = r^2$$

$$\sum_{k=0}^{M-1} c_k^2 = \sum_{k=0}^{M-1}(a_k^2 - b_k^2 + 2\mathrm{i}\,a_k b_k) = r^2\beta\exp(2\mathrm{i}\theta), \tag{86}$$

as well as the parameter

$$\eta = \frac{|P(\zeta)|}{r\sqrt{M}}, \qquad 0 < \eta \le 1. \tag{87}$$

This factor η is equal to 1 for the DFT algorithm, and is rarely less than 0.8 for almost all published ones, as can be seen in the table in the Appendix. From all the preceding definitions and results, it is easily shown [40] that the variance of the phase noise is then given by

$$\left\langle \Delta\,\phi^2 \right\rangle = \sigma_\phi^2 = \frac{1}{M\eta^2 \mathrm{SNR}^2}\left\{1 - \beta\cos\left[2(\phi-\theta)\right]\right\}. \tag{88}$$

The sinusoidal modulation can usually be neglected, as the parameter β is often less than 0.1, and the following simplified formula can be used:

$$\sigma_\phi^2 = \frac{1}{M\eta^2 \mathrm{SNR}^2}\ . \tag{89}$$

This formula can be easily analyzed. First, it is normal that the phase variance decreases with the square of the SNR. Second, the factor M is not surprising, as the detection involves the weighted addition of M statistically independent terms. Then the significance of the factor η becomes clear: it characterizes the effect of the weighting terms. In other words, it will measure how efficiently the statistically independent intensities are used to calculate the phase. In some way, and referring to the phasor interpretation presented in Sect. 2.2.7, the factor η characterizes the geometry of the phasor construction of the first Fourier coefficient of the sum $S(\phi)$. This factor is 1 for the DFT algorithm, where collinear phasors of equal length add together (Fig. 4). For the WDFT algorithm with $\delta = \pi/3$, $\eta = 0.898$ the phasor construction is still made of collinear phasors, but of unequal lengths (Fig. 5). The third example given in Sect. 2.2.7 corresponds to $\eta = 0.497$. It can be seen in Fig. 6 that in this case the detection phasor is built in a rather inefficient way. So, this parameter can be called the *efficiency factor* and is a crucial feature of the algorithm.

3 Phase Unwrapping

After the phase is detected, it has to be unwrapped to map the physical quantity (length, displacement, etc.) which is being measured. Unwrapping means removing the 2π-jumps present in the data obtained after the phase detection process (e.g. see Fig. 1b, by adding an adequate multiple of 2π where necessary. The procedure is trivial in the one dimensional case, as it suffices to scan the data from left to right looking for a phase difference between adjacent pixels greater than a prescribed value, typically π, and to remove them by adding (or subtracting, depending on the direction of the jump) 2π to the data set on the right-hand side of the jump. This path-oriented procedure often fails in the two-dimensional case because of the noise present in the data, which causes false jumps to be detected or real jumps to be missed. The essential point is that the error propagates as the phase unwrapping procedure is going on, and a result such as that in Fig. 12 may be obtained.

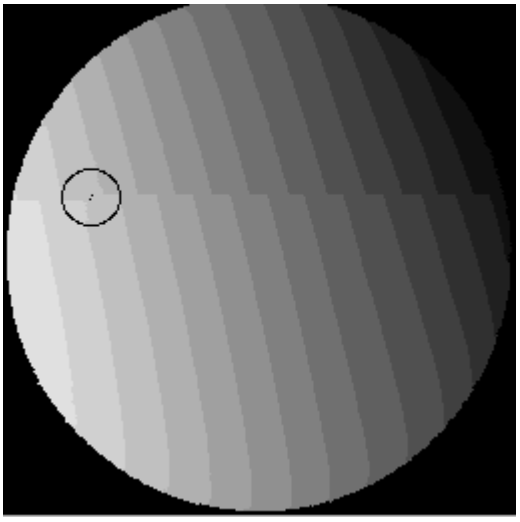


Fig. 12. Propagation of the phase unwrapping error (from the same data set as in Fig. 1). The *circle* is centered at the faulty noisy pixel

Two main strategies exist for phase unwrapping, namely spatial and temporal. However, it should be noted that the temporal one is not always possible.

3.1 Spatial Approach

Phase unwrapping is called spatial when a single wrapped phase map is used. Spatial information and operations are used to carry out the unwrapping. In this case, the problem may be quite odd, as it sometimes happens to have no solution, as in the situation sketched in Fig. 13. In Fig. 13, there are three jumps from P to Q along path A, and only two along path B. Such situations occur in speckle fringe patterns when dislocations are present in the wavefront. Points where the electromagnetic field has zero amplitude have no defined phase, and the wavefront can be skewed around those points [42].

Among the path-dependent methods, a solution is to select from among all possible paths by using amplitude [43] or phase gradient [44,45] information to increase confidence in the unwrapping path. A statistical approach for path selection using genetic algorithms have also been proposed recently [46].

Another possibility is to introduce cuts in the phase map to forbid erroneous paths[1]. It is quite easy to detect points where cuts have to end (*Huntley* and *Buckland* [47] call them "monopoles") by local unwrapping following a 3×3 pixels square around every pixel in the wrapped phase map. When the

[1] This is very similar to what is sometimes necessary when integrating a function of a complex variable

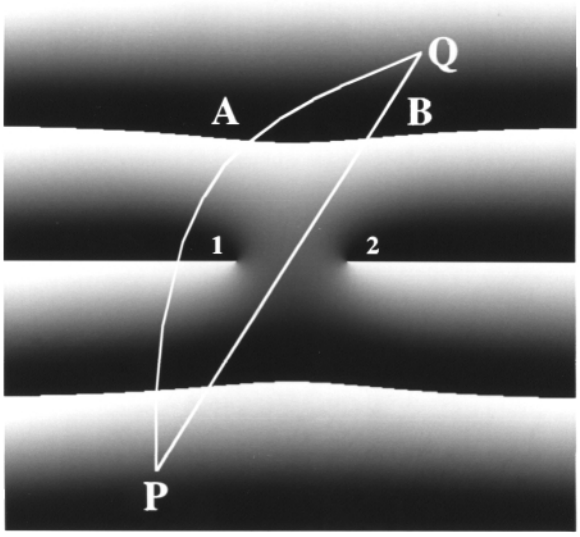


Fig. 13. Example of a phase map that cannot be unwrapped. A cut is necessary between points 1 and 2 to forbid path B (from [41])

phase obtained after looping is the same as the one at the beginning of the closed path, the central pixel is not a monopole. When it differs by $\pm 2\pi$, it is a positive or negative monopole. Once all monopoles have been identified, pairs of opposite signs must be selected to be the ends of the cuts. A simple strategy to pair monopoles is to minimize the total length of the cuts, that is to minimize the total width of the dipoles, with special attention paid to the boundaries [48,49,50].

It is also possible to segment the phase data into areas between fringe boundaries and then to calculate the phase offset between the disjoint areas [51], or into smaller regions ("tiles") which are then connected according to the confidence attached to the consistency of the phase data at the edges of the tiles [44].

Truly path-independent methods are based on the equivalence between the phase unwrapping problem and the resolution of Poisson's equation with Neumann boundary conditions, which can often be completed using specialized equation solvers. This efficient approach can be found in the literature [52,53,54], but will not be detailed here because of the level of mathematics involved.

Before closing this topic, it is worth mentioning a very simple approach which may be satisfactory for a large number of practical cases. One must realize that spatial smoothing is a local process that can be "2π-jumps aware", that is, the smoothing algorithm can perform local phase unwrapping before smoothing, and then wrap again the phase. This is illustrated by Fig. 14a,b.

The difference between the raw and smoothed data mostly represents noise (and maybe some signal) and can be retained in a buffer. The smoothed data is then unwrapped using a basic path-dependent unwrapping algorithm, as in Fig. 14c, and the previous phase difference is added back to obtain the unwrapped phase exactly corresponding to the initial wrapped one, as in Fig. 14d.

3.2 Temporal Approach

Temporal phase unwrapping is possible when the frequency of the fringes can be arbitrarily varied. It has been proposed mainly in the domain of profilometry by fringe projection. The principle of profilometry [4] is recalled in Fig. 15. Fringes (or parallel illuminated lines) are obliquely projected onto the object under investigation. The height information z from an arbitrary reference plane surface is encoded into a lateral displacement u_x of the fringe, with

$$z = \frac{u_x}{\tan\theta} \ , \tag{90}$$

in the simplest case of a collimated illumination shown in Fig. 15a. In (90) it is assumed that the direction of the fringes is aligned with the y-axis. The intensity which can be recorded by a video camera is written as

$$I(x,y) = I_0(x,y)\left\{1 + \gamma(x,y)\cos\left[\frac{2\pi x}{p} + \psi(x,y)\right]\right\} \ , \tag{91}$$

where $I_0(x,y)$ and $\gamma(x,y)$ are the slowly varying bias and contrast, p is the initial fringe pitch and $\psi(x,y)$ is the phase modulation caused by the height variation:

$$\psi(x,y) = \frac{2\pi u_x}{p} = \frac{2\pi z \tan\theta}{p} \ . \tag{92}$$

So, the height is known when the unwrapped phase is known. Two specific features of this technique are the possibility of an arbitrary variation of the projected fringe density (usually a slide or video projector is used) and the possible presence of non connected objects or stepwise height variations corresponding to phase steps much greater than 2π, making any spatial phase unwrapping method fail. So the concept of temporal phase unwrapping has emerged. In this approach extra information is needed, thus requiring more images to be recorded.

3.2.1 Gray Code

In this approach, the fringe order is encoded with a succession of 0's and 1's by the use of projected binary masks, the pitch of which varies as successive

(**a**)

(**b**)

(**c**)

(**d**)

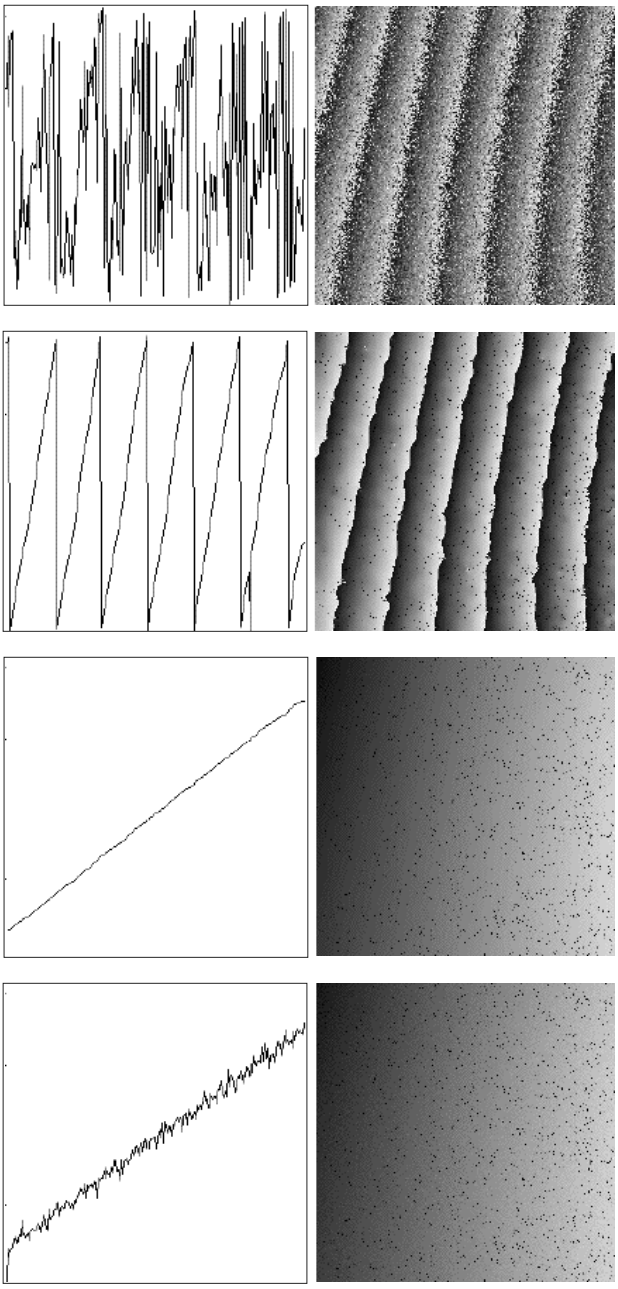


Fig. 14. Phase unwrapping by spatial smoothing: (**a**) noisy wrapped phase, (**b**) spatially smoothed phase, (**c**) smoothed unwrapped phase, and (**d**) noisy unwrapped phase

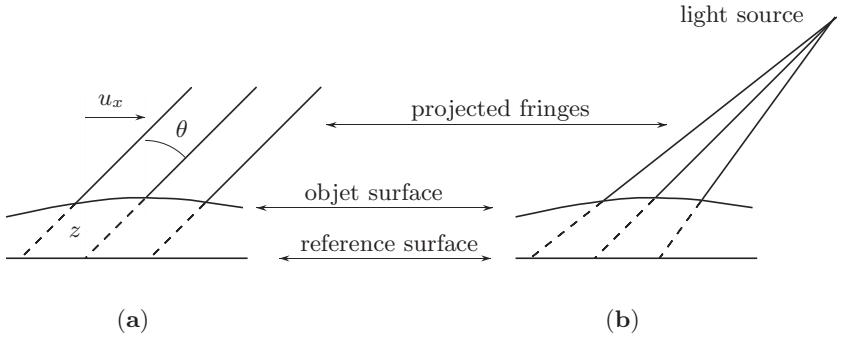


Fig. 15. Principle of projected fringes profilometry: (**a**) collimated projection, (**b**) conical projection

powers of 2. The process is illustrated in Fig. 16. This procedure is used in commercial systems like ATOS from Newport or the products from OMECA Messtechnik. The total number of frames to acquire is the number required by temporal phase-stepping plus the number of masks to project. This number is $\mathrm{INT}(\log_2 F) + 1$, where F is the total number of fringes in the field and INT is the 'integer part' function. For example, if the temporal phase-stepping uses five frames and if $F = 240$, 13 frames $(5 + 8)$ are necessary.

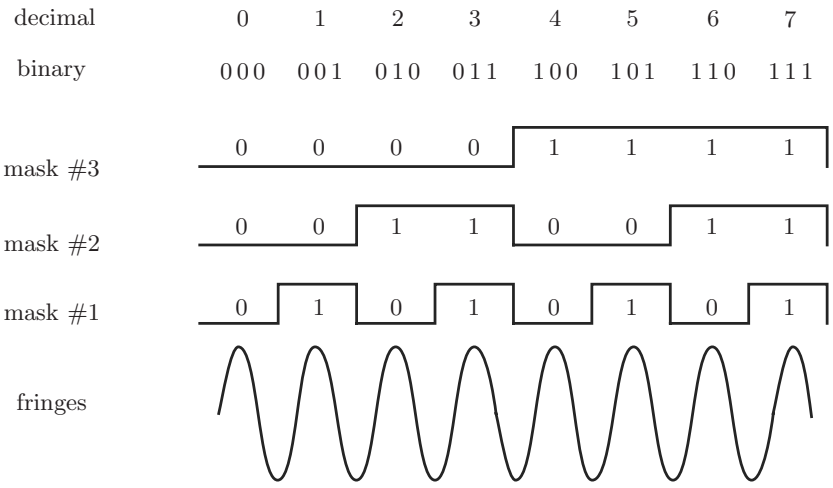


Fig. 16. Encoding of the fringe order by projection of successive binary masks

3.2.2 Multistep Process

In what we call the multistep process, many images with different fringe frequencies are used. In [41,55] the total number of projected fringes is varied over the field of view, one by one from 1 to F, with the central fringe remaining fixed. For each fringe density, a four-frame temporal phase-stepping is done, and the unwrapped phase $\phi(x, y)$ is obtained as the sum of the differences $\varphi_{i+1}(x, y) - \varphi_i(x, y)$ between the phases detected during steps i and $i + 1$. The fact that the variation of the fringe number over the field of view is no more than 1 and that the central fringe remains steady ensures that this phase difference lies within the range $[-\pi, \pi]$. The overall procedure is sketched in Fig. 17. As can be seen, the total number of frames required to carry out the processing is $4F$. However, strategies exist to decrease this number [56].

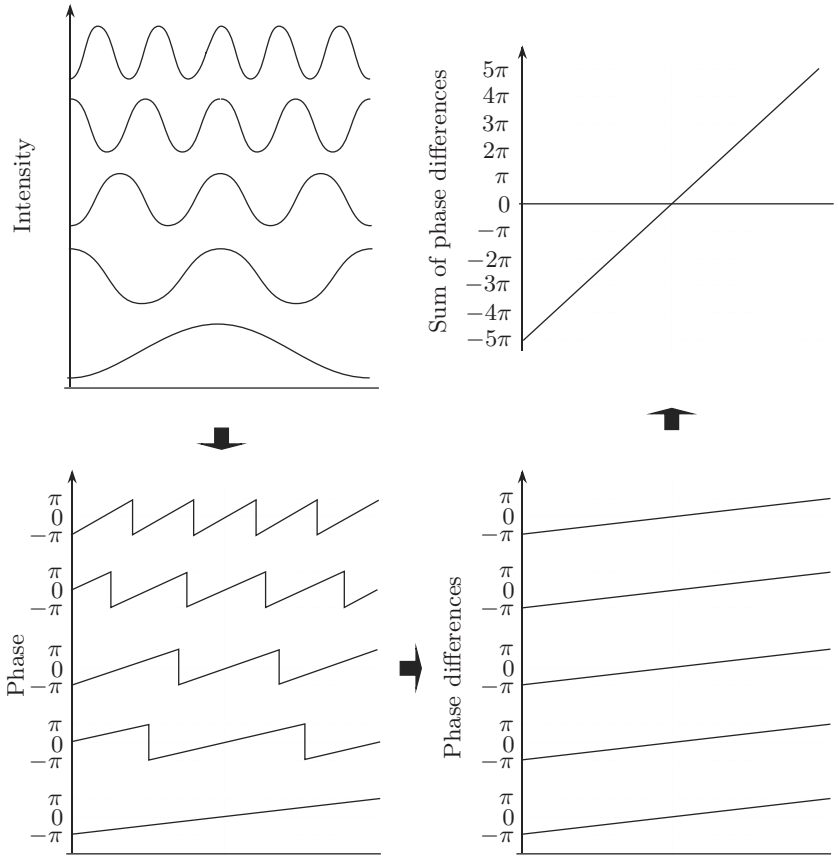


Fig. 17. Multistep temporal phase unwrapping

Recently, *Li* [57] have proposed a method where both a high- and a low-frequency grating are used, with an integer ratio between the frequencies. The high-frequency one provides the sensitivity, and the low-frequency one information for unwrapping. The two gratings are projected simultaneously, and a nice temporal phase-stepping scheme allow to implement an algorithm to sort out the low- and high-frequency phases. The total number of frames required by this method is $2P$, where P is the number of 2π-jumps that can be removed. Actually, the temporal phase-stepping is composed of two sets of P shifts, each set being necessary to deal with one of the frequencies. So, six frames is the minimum number of frames necessary to implement the method, as three frames is the minimum number of steps for the phase detection. With so few frames, only three 2π-jumps can be removed.

3.2.3 Two-Step Process

A two-step method has been proposed in [58], and a very similar method can be found in [59]. To carry out the unwrapping procedure, two fringe patterns with slightly different pitches are projected onto the object, and two images are taken, from which two slightly different phase maps are computed. The difference gives the low-frequency information enabling us to unwrap the phase. This looks like the two-frequency method by *Li* [57] indicated in the preceding section. The difference is twofold: spatial instead of temporal phase-stepping is used, and two high-frequency patterns are used.

More precisely, the slowly varying phase map obtained by subtraction is used to number the fringes. The pitch of the grating is varied between the two images by $(L-1)/L$, where L is an integer. So, the phase difference between the two computed phase maps varies by 2π for every L fringes. This phase difference can be called the "numbering phase". This numbering phase divided by $2\pi/L$ varies by L every L fringes. The integer part of this quotient straightforwardly gives the fringe order. However, the noise present in the data makes this process tail in the vicinity of the 2π-jumps, where the phase and the computed fringe order are "unstable". It is very easy to circumvent this problem. First, the preceding procedure is restricted to the half-fringes where the wrapped phase is in the interval $[-\pi/2, \pi/2]$. Then, π is substracted from the wrapped phase and π/L from the numbering phase, so that the 2π-jumps and the fringe-order transition points move out of the regions concerned. Then, the half fringes where the wrapped phase is in the interval $[-\pi, -\pi/2] \cup [\pi/2, \pi]$ can be processed. This procedure is presented in Fig. 18. However, two points are critical in this procedure: the precision of the pitch variation and the coincidence of the zero phases in the two phase maps at the central point. Obviously, the proposed procedure can unwrap up to L 2π-jumps. As L can be quite large, the dynamic range of phase detection is fairly enlarged.

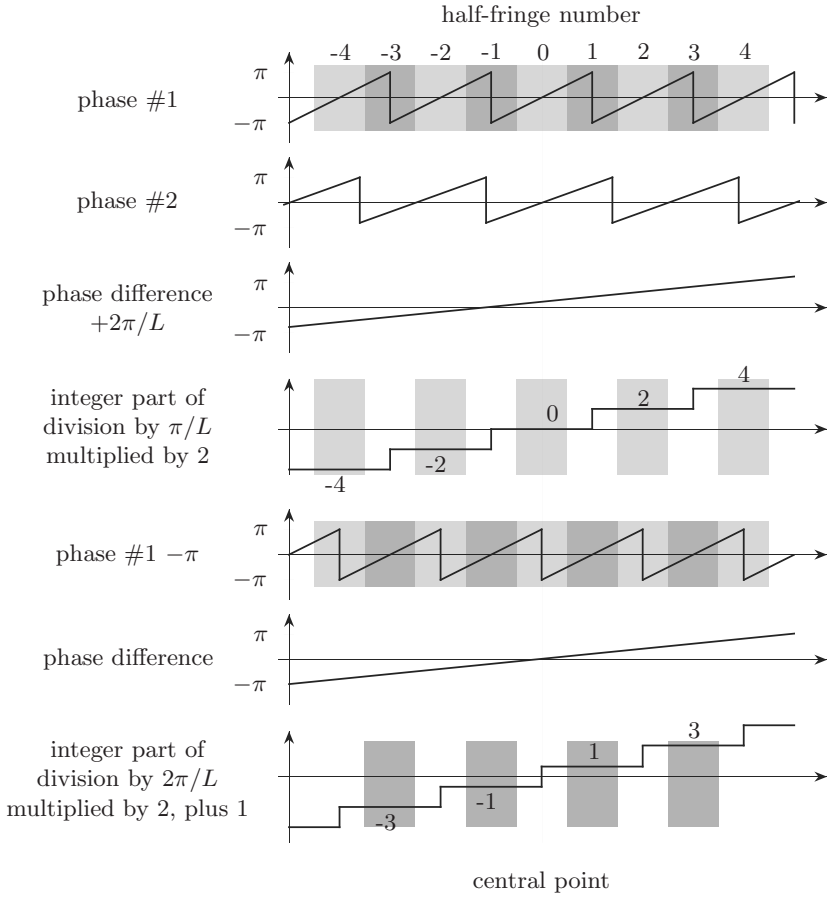


Fig. 18. Two-step temporal phase unwrapping

4 Conclusion

Tremendous activity has been observed in the field of optical metrology during the last decade. The possibility of easy and cheap image processing has completely modified the available optical tools. Methods that were most by more than one century old have been revised and their use is now spreading into industry.

Actually, it is necessary to assess the metrological performance that is required by industrial applications. Issues such as spatial resolution, sensibility and accuracy have to be addressed by following a well-defined methodology. The first step is the assessment of the phase detection from intensity fields. For the phase-stepping process, the influence of the noise and the amplitude of the residual errors can be evaluated. The same level of achievement has not yet obtained for the Fourier transform method.

The CP theory allows us to readily analyze the properties of the numerous algorithms that have been published over the years. For practical applications, only two algorithms need be considered in the first instance, namely the DFT (or N-bucket) and the WDFT ones, the parameters of which have to be determined as a function of harmonic content, amount of miscalibration and possible number of frames.

Phase unwrapping remains a difficult problem to handle in the most general case, and so remains difficult to fully automate. However, a number of strategies exist that will allow us to successfully unwrap the phase in most practical situations.

Appendix: Table of Algorithms

In Table 1, we give information about a large number of algorithms that have been published. We indicate the total number M of frames in the algorithm, the nominal phase shift δ and the characteristic diagram. The coefficients of the algorithm

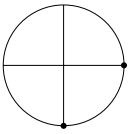
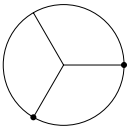
$$\tilde{\phi} = \arctan\left(\frac{\sum_{k=0}^{M-1} b_k I_k}{\sum_{k=0}^{M-1} a_k I_k}\right),$$

are written as

$$\frac{b_0, b_1, \ldots, b_{M-1}}{a_0, a_1, \ldots, a_{M-1}}.$$

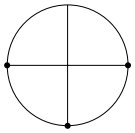
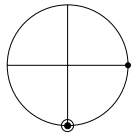
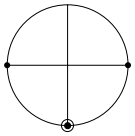
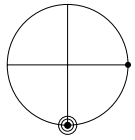
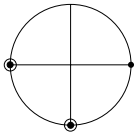
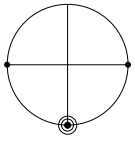
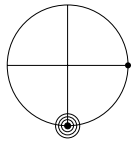
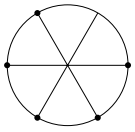
Following (15) in Sect. 2.2.2, each algorithm detects $\phi + \psi$, where $\psi = \arg[P(\zeta)]$ is a constant that is given in the Table. The sensitivity to noise propagation depends on the efficiency parameter η and on the parameter β that appear in (88) giving the standard deviation of the phase noise in Sect. 2.5.2.

Table 1. Algorithms that can be found in the literature

#	M	δ	Diagram	Coefficients	ψ	η	β	Ref.
1	3	$\frac{\pi}{2}$		$\frac{1,-1,0}{0,1,-1}$ $\frac{1,-2,1}{1,0,-1}$	$\frac{\pi}{4}$ 0	0.816	0.500	[12,60]
2	3	$\frac{2\pi}{3}$		$\frac{\sqrt{3}[0,-1,1]}{[2,-1,-1]}$	0	1	0	[12]

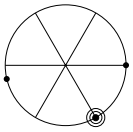
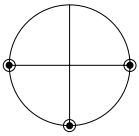
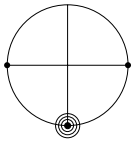
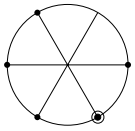
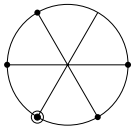
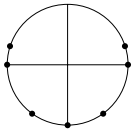
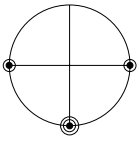
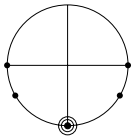
continued next page

Table 1. *(continued)*

#	M	δ	Diagram	Coefficients	ψ	η	β	Ref.
3	4	$\frac{\pi}{2}$		$\frac{1,1,-1,-1}{-1,1,1,-1}$ $\frac{0,-1,0,1}{1,0,-1,0}$	$\frac{3\pi}{4}$ 0	1	0	[61]
4	4	$\frac{\pi}{2}$		$\frac{0,-2,2,0}{1,-1,-1,1}$	0	0.816	0.333	[62,26]
5	5	$\frac{\pi}{2}$		$\frac{0,-2,0,2,0}{1,0,-2,0,1}$	0	0.956	0.143	[63,26]
6	5	$\frac{\pi}{2}$		$\frac{1,-4,0,4,-1}{1,2,-6,2,1}$	0	0.800	0.150	[26]
7	6	$\frac{\pi}{2}$		$\frac{0,-2,-2,2,2,0}{1,1,-2,-2,1,1}$	$\frac{\pi}{4}$	0.873	0.143	[64]
8	6	$\frac{\pi}{2}$		$\frac{0,-3,0,4,0,-1}{1,0,-4,0,3,0}$	0	0.906	0	[26]
9	6	$\frac{\pi}{2}$		$\frac{0,-4,4,4,-4,0}{1,-1,-6,6,1,-1}$	$-\frac{\pi}{4}$	0.781	0.086	[26]
10	6	$\frac{\pi}{3}$		$\frac{\sqrt{3}[-1,-2,-1,1,2,1]}{3,0,-3,-3,0,3}$	$-\frac{\pi}{6}$	1	0	

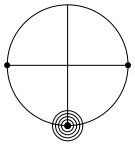
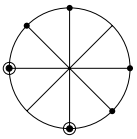
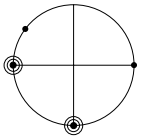
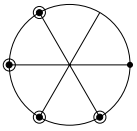
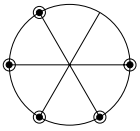
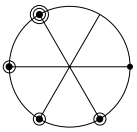
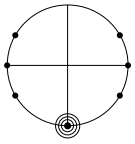
continued next page

Table 1. *(continued)*

#	M	δ	Diagram	Coefficients	ψ	η	β	Ref.
11	6	$\frac{\pi}{3}$		$\frac{\sqrt{3}[-5,6,17,-17,-6,5]}{1,-26,25,25,-26,1}$	$\frac{5\pi}{6}$	0.495	0.107	[33]
12	7	$\frac{\pi}{2}$		$\frac{1,0,-3,0,3,0,-1}{0,2,0,-4,0,2,0}$	$\frac{\pi}{2}$	0.912	0.091	[65,25]
13	7	$\frac{\pi}{2}$		$\frac{-1,0,7,0,-7,0,1}{0,-4,0,8,0,-4,0}$	π	0.864	0.020	[28]
14	7	$\frac{\pi}{3}$		$\frac{\sqrt{3}[-1,3,3,0,-3,-3,1]}{3[-1,-1,1,2,1,-1,-1]}$	π	0.953	0.118	[34]
15	7	$\frac{\pi}{3}$		$\frac{\sqrt{3}[0,1,1,0,-1,-1,0]}{-1,-1,1,2,1,-1,-1}$	π	0.967	0.091	[34]
16	8	$\frac{\pi}{2}$		$\frac{-1,-2,3,4,-4,-3,2,1}{1,-2,-3,4,4,-3,-2,1}$	$-\frac{\pi}{4}$	0.913	0	[26]
17	8	$\frac{\pi}{2}$		$\frac{-1,-3,5,7,-7,-5,3,1}{1,-3,-5,7,7,-5,-3,1}$	$-\frac{\pi}{4}$	0.873	0	[26]
18	8	$\frac{\pi}{2}$		$\frac{-1,-4,8,11,-11,-8,4,1}{1,-4,-8,11,11,-8,-4,1}$	$-\frac{\pi}{4}$	0.844	0	[26]

continued next page

Table 1. *(continued)*

#	M	δ	Diagram	Coefficients	ψ	η	β	Ref.
19	8	$\frac{\pi}{2}$		$\frac{-1,-5,11,15,-15,-11,5,1}{1,-5,-11,15,15,-11,-5,1}$	$-\frac{\pi}{4}$	0.830	0	[26]
20	9	$\frac{\pi}{4}$		$\frac{0,1,2,1,0,-1,-2,-1,0}{-1,-1,0,1,2,1,0,-1,-1}$	π	0.970	0.091	[28]
21	9	$\frac{\pi}{2}$		$\frac{1,-2,-14,-18,0,18,14,2,-1}{2,8,8,-8,-20,-8,8,8,2}$	π	0.497	0.140	[33]
22	10	$\frac{\pi}{3}$		$\frac{\sqrt{3}[1,-1,-7,-11,-6,6,11,7,1,-1]}{3[1,3,3,-1,-6,-6,-1,3,3,1]}$	$\frac{\pi}{2}$	0.830	0.106	[25]
23	11	$\frac{\pi}{3}$		$\frac{\sqrt{3}[-1,-2,0,4,5,0,-5,-4,0,2,1]}{1,-2,-6,-4,5,12,5,-4,-6,-2,1}$	$-\frac{\pi}{3}$	0.898	0.055	[40]
24	11	$\frac{\pi}{3}$		$\frac{\sqrt{3}[0,1,4,7,6,0,-6,-7,-4,-1,0]}{-2,-5,-6,-1,8,12,8,-1,-6,-5,-2}$	$-\frac{\pi}{3}$	0.681	0.205	[65]
25	11	$\frac{\pi}{2}$		$\frac{1,0,-8,0,15,0,-15,0,8,0,-1}{0,4,0,-12,0,16,0,-12,0,4,0}$	$\frac{\pi}{2}$	0.851	0.003	[28]

References

1. F. Roddier, C. Roddier: Imaging with a Multi-Mirror Telescope, in F. Pacini, W. Richter, R. N. Wilson (Ed.): *ESO Conference on Optical Telescopes of the Future, Geneva*, ESO, CERN, Geneva (1978) 57, 77
2. C. Roddier, F. Roddier: Imaging with a Coherence Interferometer in Optical Astronomy, in C. van Schooneveld (Ed.): *IAU Colloquium 49 on Formation of Images from Spatial Coherence Functions in Astronomy, Groningen, 1978*. Reidel, Norwell, MA (1979) 57, 77
3. M. Takeda, H. Ina, S. Kobayashi: Fourier-Transform Method of Fringe-Pattern Analysis for Computer-Based Topography and Interferometry, J. Opt. Soc. Am. A. **72**, 156–160 (1982) 57, 77
4. M. Takeda, K. Mutoh: Fourier Transform Profilometry for the Automatic Measurement of 3-D Object Shapes, Appl. Opt. **22**, 3977–3982 (1983) 57, 77, 89
5. W. W. Macy, Jr.: Two-Dimensional Fringe-Pattern Analysis, Appl. Opt. **22**, 3898–3901 (1983) 57, 77
6. K. H. Womack: Frequency Domain Description of Interferogram Analysis, Opt. Eng. **23**, 396–400 (1984) 57, 77
7. M. Takeda: Spatial Carrier Fringe Pattern Analysis and its Applications to Precision Interferometry and Profilometry: An Overview, Ind. Metrol. **1**, 79–99 (1990) 57, 77
8. J. B. Liu, P. D. Ronney: Modified Fourier Transform Method for Interferogram Fringe Pattern Analysis, Appl. Opt. **36**, 6231–6241 (1997) 57, 77
9. P. Carré: Installation et Utilisation du Comparateur Photoélectrique et Interférentiel du Bureau International des Poids et Mesures, Metrologia **2**, 13–23 (1966) 57
10. J. H. Bruning, D. R. Herriott, J. E. Gallagher, D. P. Rosenfeld, A. D. White, D. J. Brangaccio: Digital Wavefront Measuring Interferometer for Testing Optical Surfaces and Lenses, Appl. Opt. **13**, 2693–2703 (1974) 57
11. J. C. Wyant: Use of an AC Heterodyne Lateral Shear Interferometer with Real Time Wavefront Correction Systems, Appl. Opt. **14**, 2622–2626 (1975) 57
12. C. L. Koliopoulos: Interferometric Optical Phase Measurement Techniques, PhD thesis, University of Arizona, Tucson, AZ (1981) 57, 63, 95, 95
13. K. A. Stetson, W. R. Brohinsky: Electrooptic Holography and its Application to Hologram Interferometry, Appl. Opt. **24**, 3631–3637 (1985) 57
14. J. H. Bruning: Fringe Scanning Interferometers, in D. Malacara (Ed.): *Optical Shop Testing*, Wiley, New York (1987), 414 57
15. K. H. Womack: Interferometric Phase Measurement Using Spatial Synchronous Detection, Opt. Eng. **23**, 391–395 (1984) 58
16. S. Toyooka, M. Tominaga: Spatial Fringe Scanning for Optical Phase Measurement, Opt. Commun. **51**, 68–70 (1984) 58
17. M. Kujawińska, J. Wójciak: Spatial-Carrier Phase-Shifting Technique of Fringe Pattern Analysis, in W. P. Jüptner (Ed.): *Industrial Applications of Holographic and Speckle Measuring Techniques*, Proc. SPIE **1508**, 61–67 (1991) 58
18. X. Colonna de Lega: Processing of Non-stationary Interference Patterns: Adapted Phase-Shifting Algorithms and Wavelet Analysis. Application to Dynamic Deformation Measurements by Holographic and Speckle Interferometry, PhD thesis, EPFL, Lausanne, Switzerland (1997) 58

19. J. L. Maroquin, M. Servin, R. Rodriguez-Vera: Adaptive Quadrature Filters and the Recovery of Phase from Fringe Pattern Images, Appl. Opt. **14**, 1742–1753 (1997) 58
20. J. Schwider, R. Burow, K.-E. Elssner, J. Grzanna, R. Spolaczyk, K. Merkel: Digital Wave-Front Measuring Interferometry: Some Systematic Error Sources, Appl. Opt. **22**, 3421–3432 (1983) 58, 73, 73, 74
21. C. Ai, J. C. Wyant: Effect of Piezoelectric Transducer Nonlinearity on Phase Shift Interferometry, Appl. Opt. **26**, 1112–1116 (1987) 58
22. J. van Wingerden, H. J. Frankena, C. Smorenburg: Linear Approximation for Measurement Errors in Phase-Shifting Interferometry, Appl. Opt. **30**, 2718–2729 (1991) 58
23. K. Creath: Phase Measurement Interferometry: beware these Errors, in R. J. Pryputniewicz (Ed.): *Laser Interferometry IV: Computer-Aided Interferometry*, SPIE **1553**, 213–220 (1992) 58
24. K. Freischlad, C. L. Koliopoulos: Fourier Description of Digital Phase-Measuring Interferometry, J. Opt. Soc. Am. A. **7**, 542–551 (1990) 59, 71, 72
25. Y. Surrel: Design of Algorithms for Phase Measurements by the Use of Phase-Stepping, Appl. Opt. **35**, 51–60 (1996) 59, 62, 64, 83, 83, 97, 98
26. J. Schmit, K. Creath: Extended Averaging Technique for Derivation of Error-Compensating Algorithms in Phase-Shifting Interferometry, Appl. Opt. **34**, 3610–3619 (1995) 59, 73, 73, 96, 96, 96, 96, 96, 97, 97, 97, 98
27. J. Schmit, K. Creath: Window Function Influence on Phase Error in Phase-Shifting Algorithms, Appl. Opt. **35**, 5642–5649 (1996) 59, 73, 74, 75
28. P. de Groot: Derivation of Algorithms for Phase-Shifting Interferometry Using the Concept of Data-Sampling Window, Appl. Opt. **34**, 4723–4730 (1995) 59, 73, 74, 97, 98, 98
29. D. W. Phillion: General Methods for Generating Phase-Shifting Interferometry Algorithms, Appl. Opt. **36**, 8098–8115 (1997) 59, 75, 75, 76, 76
30. R. Onodera, Y. Ishii: Phase-Extraction Analysis of Laser-Diode Phase-Shifting Interferometry that is Insensitive to Changes in Laser Power, J. Opt. Soc. Am. A. **13**, 139–146 (1996) 66
31. Y. Surrel: Design of Phase-Detection Algorithms Insensitive to Bias Modulation, Appl. Opt. **36**, 805–807 (1997) 66, 66
32. Y. Surrel: Phase-Shifting Algorithms for Nonlinear and Spatially Nonuniform Phase Shifts: Comment, J. Opt. Soc. Am. A. **15**, 1227–1233 (1998) 67, 83
33. K. Hibino, B. F. Oreb, D. I. Farrant, K. G. Larkin: Phase-Shifting Algorithms for Nonlinear and Spatially Nonuniform Phase Shifts, J. Opt. Soc. Am. A. **14**, 918–930 (1997) 68, 97, 98
34. K. G. Larkin, B. F. Oreb: Design and Assessment of Symmetrical Phase-Shifting Algorithms, J. Opt. Soc. Am. A. **9**, 1740–1748 (1992) 71, 97, 97
35. P. de Groot: 101-Frame Algorithm for Phase-Shifting Interferometry, in: *Optical Inspection and Micromeasurements II*, Proc. SPIE **3098**, 283–292 (1997) 73
36. Y. Surrel: Extended Averaging and Data Windowing Techniques in Phase-Stepping Measurements: An Approach Using the Characteristic Polynomial Theory, Opt. Eng. **37**, 2314–2319 (1998) 75
37. D. J. Bone, H. A. Bachor, R. John Sandeman: Fringe-pattern Analysis Using a 2-D Fourier Transform, Appl. Opt. **25**, 1653–1660 (1986) 78

38. H. T. Goldrein, S. J. P. Palmer, J. M. Huntley: Automated Fine Grid Technique for Measurement of Large-Strain Deformation Map, Opt. Lasers Eng. **23**, 305–318 (1995) 78
39. Y. Surrel: Phase Stepping: A New Self-Calibrating Algorithm, Appl. Opt. **32**, 3598–3600 (1993) 83
40. Y. Surrel: Additive Noise Effect in Digital Phase Detection, Appl. Opt. **36**, 271–276 (1997) 84, 84, 86, 98
41. J. M. Huntley, H. O. Saldner: Temporal Phase-Unwrapping Algorithm for Automated Interferogram Analysis, Appl. Opt. **32**, 3047–3052 (1993) 88, 92
42. N. B. Baranova, B. Ya. Zel'dovich: Dislocations of the Wave-front Surface and Zeros of the Amplitude, Sov. Phys.-JETP **30**, 925–929 (1981) 87
43. M. Takeda, T. Abe: Phase Unwrapping by a Maximum Cross-Amplitude Spanning Tree Algorithm: A Comparative Study, Opt. Eng. **35**, 2345–2351 (1996) 87
44. D. P. Towers, T. R. Judge, P. J. Bryanston-Cross: Automatic Interferogram Analysis Techniques Applied to Quasi-heterodyne Holography and ESPI, Opt. Lasers Eng. **14**, 239–282 (1991) 87, 88
45. T. R. Judge, C. Quan, P. J. Bryanston-Cross: Holographic Deformation Measurements by Fourier Transform Technique with Automatic Phase Unwrapping, Opt. Eng. **31**, 533–543 (1992) 87
46. A. Collaro, G. Franceschetti, F. Palmieri, M. S. Ferreiro: Phase Unwrapping by Means of Genetic Algorithms, J. Opt. Soc. Am. A **15**, 407–418 (1998) 87
47. J. M. Huntley, J. R. Buckland: Characterization of Sources of 2π Phase Discontinuity in Speckle Interferograms, J. Opt. Soc. Am. A. **12**, 1990–1996 (1995) 87
48. R. Cusack, J. M. Huntley, H. T. Goldrein: Improved Noise-Immune Phase-Unwrapping Algorithm, Appl. Opt. **35**, 781–789 (1995) 88
49. J. R. Buckland, J. M. Huntley, S. R. E. Turner: Unwrapping Noisy Phase Maps by Use of a Minimum-Cost-Matching Algorithm, Appl. Opt. **34**, 5100–5108 (1995) 88
50. J. A. Quiroga, A. González-Cano, E. Bernabeu: Stable-Marriages Algorithm for Preprocessing Phase Maps with Discontinuity Sources, Appl. Opt. **34**, 5029–5038 (1995) 88
51. P. G. Charette, I. W. Hunter: Robust Phase-unwrapping Method for Phase Images with High Noise Content, Appl. Opt. **35**, 3506–3513 (1996) 88
52. D. C. Ghiglia, L. A. Romero: Robust Two-Dimensional Weighted and Unweighted Phase Unwrapping that Uses Fast Transforms and Iterative Methods, J. Opt. Soc. Am. A. **11**, 107–117 (1994) 88
53. D. C. Ghiglia, L. A. Romero: Minimum L^p-Norm Two-Dimensional Phase Unwrapping, J. Opt. Soc. Am. A. **13**, 1999–2013 (1996) 88
54. G. H. Kaufmann, G. E. Galizzi: Unwrapping of Electronic Speckle Pattern Interferometry Phase Maps: Evaluation of an Iterative Weighted Algorithm, Opt. Eng. **37**, 622–628 (1998) 88
55. H. O. Saldner, J. M. Huntley: Profilometry Using Temporal Phase Unwrapping and a Spatial Light Modulator-Based Fringe Projector, Opt. Eng. **36**, 610–615 (1997) 92
56. J. M. Huntley, H. O. Saldner: Error Reduction Methods for Shape Measurement by Temporal Phase Unwrapping, J. Opt. Soc. Am. A. **14**, 3188–3196 (1997) 92

57. J.-L. Li, H.-J. Su, X.-Y. Su: Two-Frequency Grating Used in Phase-Measuring Profilometry, Appl. Opt. **36**, 277–280 (1997) 93, 93
58. Y. Surrel: Two-Step Temporal Phase Unwrapping in Profilometry, in C. Gorecki (Ed.): *Optical Inspection and Micromeasurements II, 16–19 June 1997, Munich*, Proc. SPIE **3098**, 271–282 (1997) 93
59. M. Fujigaki, Y. Morimoto: Automated Shape Analysis for Multiple Phase Fringes by Phase-Shifting Method Using Fourier Transform, in I. M. Allison (Ed.): *Experimental Mechanics - Advances in Design, Testing and Analysis*, Balkema, Rotterdam (1998), 711–714 93
60. J. B. Hayes: Linear Methods of Computer Controlled Optical Figuring, PhD thesis, University of Arizona, Tucson, AZ (1984) 63, 95
61. W. R. C. Rowley, J. Hamon: Quelques Mesures de Dissymétrie de Profils Spectraux, R. Opt. Théor. Instrum. **42**, 519–523 (1963) 96
62. J. Schwider, O. Falkenstörfer, H. Schreiber, A. Zöller, N. Streibl: New Compensating Four-Phase Algorithm for Phase-Shift Interferometry, Opt. Eng. **32**, 1883–1885 (1993) 96
63. P. Hariharan, B. F. Oreb, T. Eiju: Digital Phase-Shifting Interferometry: A Simple Error-Compensating Phase Calculation Algorithm, Appl. Opt. **26**, 2504–2506 (1987) 96
64. B. Zhao, Y. Surrel: Phase Shifting: A Six-Step Self-Calibrating Algorithm Insensitive to the Second Harmonic in the Fringe Signal, Opt. Eng. **34**, 2821–2822 (1995) 96
65. K. Hibino, B. F. Oreb, D. I. Farrant, K. G. Larkin: Phase-Shifting for Nonsinusoidal Waveforms with Phase-Shift Errors, J. Opt. Soc. Am. A. **12**, 761–768 (1995) 97, 98