Available online at www.sciencedirect.com
ScienceDirect
C. R. Physique 8 (2007) 298–311
COMPTES RENDUS PHYSIQUE
http://france.elsevier.com/direct/COMREN/

Optical techniques for direct imaging of exoplanets/Techniques optiques pour l'imagerie directe des exoplanètes

The Four Quadrant Phase Mask Coronagraph and its avatars

Daniel Rouan [a,b,*], Jacques Baudrand [a,b], Anthony Boccaletti [a,b], Pierre Baudoz [a,b], Dimitri Mawet [c], Pierre Riaud [a,b]

[a] *Observatoire de Paris, LESIA-CNRS, 5, place Jules-Janssen, 92195 Meudon cedex, France*
[b] *Groupement d'intérêt scientifique PHASE (Partenariat Haute résolution Angulaire Sol Espace), France*
[c] *Université de Liège, institut d'astrophysique et de géophysique de Liège, B5C, B-4000 Sart Tilman, Belgium*

Available online 12 June 2007

Abstract

A renewal of interest into Lyot coronagraphy began about 10 years ago with the advent of the new and very demanding goal of detecting directly extrasolar planets. Among the new ideas, the Four Quadrant Phase Mask (4QPMC) has been rather prolific. The 4QPM coronagraph is based the peculiar design of binary phase mask $(0, \pi)$ dividing the full field of view at the focal plane in four quadrants. The mutual destructive interferences of the coherent light of a source perfectly centered on the mask, produce a total nulling within the pupil image, as is demonstrated analytically. We first recall the concept and performance of the 4QPMC. We then analyse the different limitations in space or ground-based observations from simulations. We then address the issue of device manufacturing and give the status of the actual performance in laboratory. A section reviews the various instruments under development that are equipped with this component. We conclude with a panorama of the rest of the family of the 4QPMC, which comprises several variants proposed by different teams. More specifically we address the important question of the achromatization by describing the various ideas which are currently explored to solve this problem. Among them, some genuine ones, not yet published, are presented. ***To cite this article: D. Rouan et al., C. R. Physique 8 (2007).***

Résumé

Le coronographe quatre quadrants à masque de phase et ses avatars. Un renouveau d'intérêt pour la coronagraphie de Lyot s'est fait jour il y a environ 10 ans avec l'émergence du but nouveau mais très exigeant de détecter directement les planètes extrasolaires. Parmi les nouvelles idées, celle du masque de phase à quatre quadrants (4QPMC) s'est montrée assez prolifique. Le coronagraphe 4QPM est basé sur le concept particulier d'un masque binaire de phase $(0, \pi)$ divisant le champ dans le plan focal suivant quatre quadrants. Les interférences destructives mutuelles de la lumière d'une source cohérente parfaitement centrée sur le masque produit une annulation totale dans l'image de pupille, comme on le démontre analytiquement. Nous rappelons d'abord le concept et les performances du 4QPMC. Nous analysons ensuite les différentes limitations des observations depuis l'espace ou du sol par des simulations, puis nous abordons la question de la fabrication du dispositif et donnons le statut des performances obtenues au laboratoire. Une revue est faite des instruments en cours de développement équipés de ce composant. Nous concluons avec un panorama du reste de la famille du 4QPMC qui comporte plusieurs variantes proposées par différentes équipes. Plus spécifiquement nous traitons la question importante de l'achromatisation avec une description des diverses idées qui sont actuellement explorées pour résoudre le problème. Parmi elles, quelques innovations, pas encore publiées, sont présentées. ***Pour citer cet article : D. Rouan et al., C. R. Physique 8 (2007).***

* Corresponding author at: Observatoire de Paris, LESIA-CNRS, 5, place Jules-Janssen, 92195 Meudon cedex, France.
E-mail addresses: daniel.rouan@obspm.fr (D. Rouan), Jacques.Baudrand@obspm.fr (J. Baudrand), Anthony.Boccaletti@obspm.fr (A. Boccaletti), Pierre.Baudoz@obspm.fr (P. Baudoz), Dimitri.Mawet@student.ulg.ac.be (D. Mawet), Pierre.Riaud@obspm.fr (P. Riaud).

1631-0705/$ – see front matter
doi:10.1016/j.crhy.2007.05.004

Keywords: Exoplanet; Coronagraphy; Direct detection

Mots-clés : Exoplanètes ; Coronographie ; Détection directe

1. Introduction: the need for efficient coronagraphy

The first direct images of extrasolar planets will certainly have an important media impact, but, more importantly, this new capability will bring new pieces of information on the physical parameters of these objects: albedo, orbital elements, combined information on temperature and composition, etc.; even seasonal variations could be monitored in the future. Even more important is that spectroscopy and thus true physics will become possible.

To reach this goal, high contrast imaging techniques have to make a big step in performance, because the detection of faint extended sources or point-like companions near a bright astrophysical object (e.g. star or AGN) is not an easy task and requires both a high angular resolution and a high dynamic range. To be convinced of the difficulty of the task, let us just consider the case of a standard Jupiter at 10 pc around a G2V star: it is 23 magnitudes (i.e. 1.3×10^9 times) fainter than its star in the visible and separated only by 0.5 arcsec from it. However, there is a large range of conditions that could make the life easier. For instance, a young massive (10 M_{Jup}) planet, still on the contraction phase—and thus at a rather high temperature—, orbiting a cold M star and observed at 20 μm, is *only* 4.8 magnitudes fainter than its star. Between those two cases, there is a variety of conditions that could be potentially frequent and accessible with the present state of the art. They represent the first targets that will allow astronomers to enter soon the era of the direct detection of extrasolar planets.

On the instrumentation side, there are two avenues to explore (Rouan [1]): in the visible to near-IR using ground-based large telescopes equipped with adaptive optics and from space, in the thermal IR, using interferometers and cold telescopes. In both cases, however, a coronagraph, blocking most of the light from the bright source, is mandatory. Such an instrument is a powerful tool in a variety of other astrophysical topics: white or brown dwarves companions, accretion disks, the environment of AGN. One can show that a classical Lyot coronagraph, where an opaque disk is used to block light, is unable to fulfill the requirements set by planet detection. One reason is the closest angular distance reachable: it cannot be smaller than the mask radius, i.e. a few λ/D. The dynamic is also insufficient. Since the renewal of the interest for stellar coronagraphy that followed the discovery of 51 PegB in 1995, there have been many new ideas and concepts that flourished to overcome those limits. The four quadrants phase mask coronagraph (4QPMC hereafter) described here is one of them. It is a rather novel solution that has several distinctive advantages: it can be very efficient when used with a filled pupil and it is rather easy to manufacture and to install in an imaging instrument. Our group at LESIA (Observatoire de Paris-Meudon) has been the main contributor to developments around this device during the past six years, and this article is largely based on the results we have obtained.

We shall first recall the basic principles of the 4QPMC, why it can be theoretically totally efficient, and present the results of simulations where realistic conditions, more representative of the real world, are taken into account. We shall then describe three cases of real instruments where 4QPMC devices were, or will soon be, installed to bring the capability of planet detection in a rather near-future: the VLT-NACO system, the MIRI instrument aboard JWST and the SPHERE instrument (formerly the VLT Planet Finder) to be installed on one unit of the VLT. In its basic form, the 4QPMC can be fully efficient only at one given wavelength: we review the various solution that have been proposed to overcome this limitation and to obtain a quasi-achromatic device. In the last section we are interested with the avatars or cousins of the 4QPMC, to illustrate how a first idea can generate many offspring.

2. Principle of the basic 4QPM: the 'perfect case'

The first concept of a stellar coronagraph based on phase modification was proposed by Gay and Rabbia in 1995 and published later (Baudoz et al. [2]). The principle, as opposed to the classical coronagraph of Lyot where the mask modifies the amplitude of the wavefront, is to produce a π phase shift on one of the arms of a Michelson interferometer. Shortly after, Roddier and Roddier [3] proposed the idea of a stellar coronagraph with a phase mask, but now in the focal plane. The principle of their coronagraph is to put at the focal plane a small transparent disk of diameter $\approx 0.5\lambda/D$ (D is the telescope diameter) that produces a phase shift of π. The Four Quadrant Phase Mask

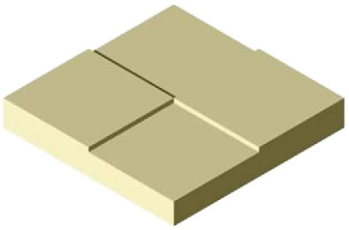

Fig. 1. Cartoon describing the structure of the FQPM device.

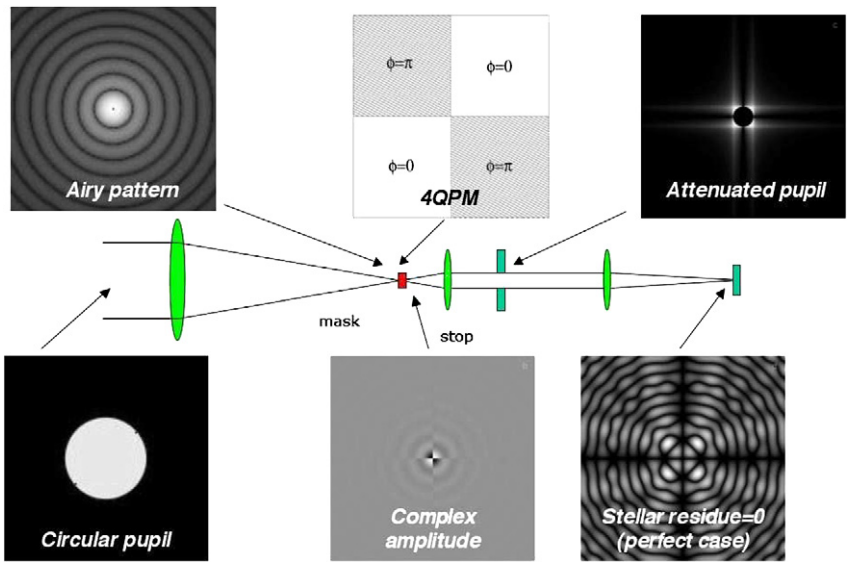

Fig. 2. Cartoon describing the optical set-up of a full FQPM coronagraphic system; the images at the various pupil or image planes are also given. The individual images are labelled. Note that images are displayed with non-linear scales.

Coronagraph that we proposed in 2000 (Rouan et al. [4]) uses as well a transmissive phase-mask in the focal plane to provide a π phase-shift on a well-defined area; however, instead of a disk, the mask, here, is arranged according to a four-quadrant pattern: two quadrants on one diagonal without phase-shift and the two other quadrants providing a π phase-shift (Fig. 1). As in the coronagraph of Bernard Lyot, the mask has to be put in a focal plane. Fig. 2 presents a layout of the optical system to be used in order to obtain the coronagraphic effect. A lens makes an image of the pupil where a Lyot stop is installed and a second lens provides the image of the focal plane. The different steps of image formation are also given in Fig. 2.

Provided that the image of a bright source is exactly located at the centre of the mask, the four beams combine in a destructive way at infinity and the stellar light is fully rejected outside of the pupil area, with a concentration of light near the edge in four spikes (see upper right of Fig. 2). A slightly undersized Lyot stop placed in this exit pupil, removes then the diffracted starlight.

Since the four-quadrant symmetry provides a perfect balancing of fluxes between each pair of quadrants, as long as the star is accurately centred, it is very efficient. In fact, the intrinsic nulling performance of the four-quadrant coronagraph is theoretically total, in the ideal case of a circular filled pupil and a perfect (aberration-free) optical system, as we show now.

In the focal plane the amplitude is described by: $a(x, y) = Airy(x, y) \times FQ(x, y)$, where $FQ(x, y) = Sign(x) \times Sign(y)$, describes the effect of the 4Q mask.

In the pupil plane the amplitude is then given by taking the Fourier transform: $w(u, v) = TF\{Airy(x, y)\} \star TF\{FQ(x, y)\}$, or $w(u, v) = Pupil(u, v) \star (uv)^{-1}$, since $TF\{Sign(x) \times Sign(y)\} = \frac{1}{uv}$.

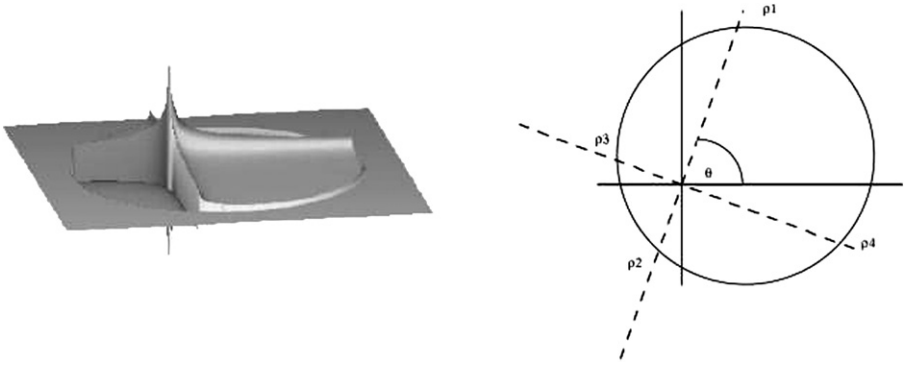

Fig. 3. Left: Shape of the $\frac{1}{uv}$ function convolved with the pupil. Right: Definition of the geometrical variables used.

For a circular pupil without any obstruction, we show now that the amplitude is indeed null at any point in the pupil, following the line of the proof proposed by Jean Gay (Observatoire de Nice). Using polar coordinates: $x = \rho\cos\theta$ and $y = \rho\sin\theta$, the last equation becomes:

$$w(u,v) = \int_0^{\pi}\left[\sin(2\theta)^{-1}\int_0^{r}\rho^{-2}\,\mathrm{d}\rho\right]\mathrm{d}\theta$$

$$w(u,v) = \int_0^{\pi}\ln\bigl[(\rho_1\times\rho_2)/(\rho_3\times\rho_4)\bigr]/\sin(2\theta)\,\mathrm{d}\theta$$

where ρ_1, ρ_2, ρ_3 and ρ_4 are described on Fig. 3. At this point, one can use a classical result of geometry which states that the product $\rho_1 \times \rho_2$ (the power of a point with respect to a circle) is a constant. This means that $[(\rho_1 \times \rho_2)/(\rho_3 \times \rho_4)] = 1$ and thus that the ln() term in the integral cancels out. As a result, here is no light at all that propagates within the geometric image of the entrance pupil!

It must be clear, however, that nulling is no longer perfect when: (a) there is a central obstruction, or if the pupil is not circular; (b) a finite wavelength bandwidth, residual optical aberrations and especially the atmospheric turbulence become important limitations, as in any of the recently proposed solutions of coronagraph; the next section will deal with this point.

Another limit is, of course, the existence of frontiers in the image plane, so that a source exactly on the frontier will be partially extinct. A faint companion lying exactly along one of the axis of the mask is attenuated by about 2.15 magnitudes owing to the π phase shift between 2 quadrants. This attenuation decreases to 0.64 magnitude when the companion is located at $0.5\lambda/D$ off one of the axis. Actually, the attenuation provided by only 2 quadrants is not very significant compared to that of a 4 quadrants phase-mask. The photometry of a companion is generally little affected this way. Moreover, to avoid this effect when looking for faint companions, a second measurement can be performed after rotating the field by 45°. The new coronagraphic concept of Annular Groove Phase Mask (Mawet et al. [5]) suppresses these dead zones (see Section 5.6).

How does the mathematical result on the total nulling compare with simulations? We performed numerical simulations of the 4QPMC, using large arrays (up to 1024×1024), so as to minimise the aliazing effect of Fast Fourier Transform and to achieve a large pixel sampling of both the pupil and image plane. Only the inner part of the Airy pattern is taken into account (up to $15\lambda/D$). Within the limits of numerical accuracy, the reality of the *perfect* nulling is fully confirmed, as well as the effect of its degradation when the star is off-axis: Fig. 4 illustrates this degradation versus angular offset from the centre.

3. The 4QPM coronagraph in the real world

Phase mask coronagraphs are indeed very sensitive to several factors: pupil shape, chromatism, wave-front distortions, residual jitter. In this section, we examine how the 4QPMC behaves in a more concrete world, first, by

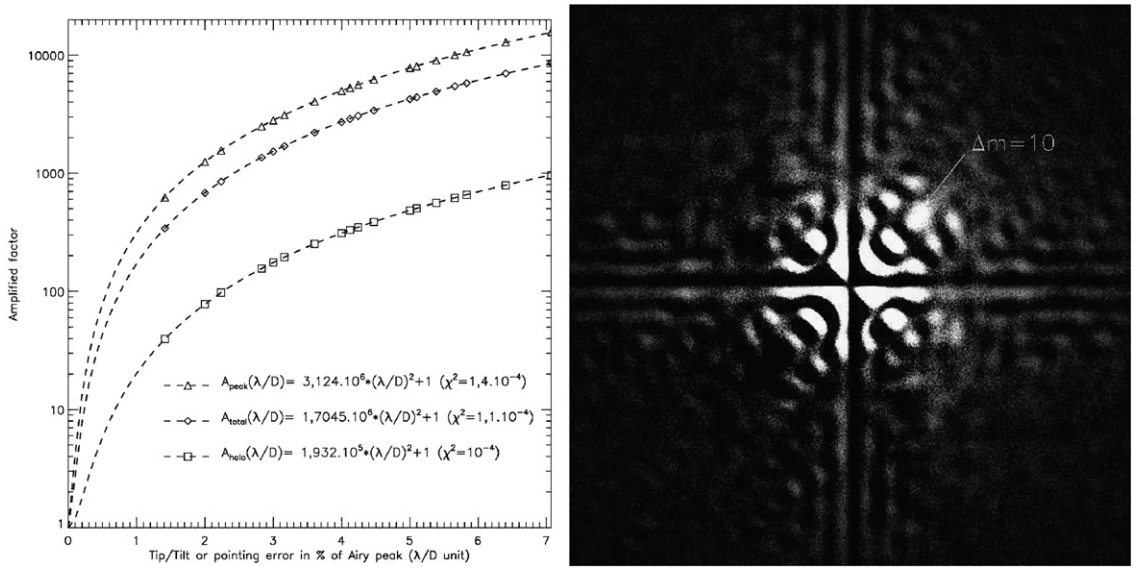

Fig. 4. Left: Numerical simulation of the four-quadrants coronagraph efficiency when the source is off-axis, in the perfect case of a filled circular pupil and no aberrations. The quantity plotted is the residual energy normalised to the perfectly centred case, as measured in the entire Airy pattern (diamonds), in the Airy rings (squares), and in the Airy peak (triangles). The offset is performed along a diagonal of the mask and a Lyot stop of 95% of the pupil diameter is used. Right: Numerical simulation in the case of realistic conditions: a median Strehl ratio of 0.79 ± 0.2 including a 25% central obscuration and spider arms. Photon noise (total integration of 1 hour on an $m_K = 4.9$ star with a 3.6 m telescope) and read-out noise (5 e- pixel-1 frame-1) are added to each of the 2000 individual frames. A companion 8 mag fainter than the central star is detected at a distance of $3\lambda/D$ after subtraction of the opposite quadrants.

simulating the nulling performance of the device when there are aberrations of the wavefront and especially tip/tilt residuals; second, by considering the question of manufacturing the component and finally by looking at performances at the laboratory.

3.1. Performance estimation of a real system

To define a performance, it is necessary to define a metric. We consider three different ones:

- Total rejection: Ratio of total intensity of the direct image to that of the coronagraphic image. It only gives the global attenuation of the on-axis object but can be easily compared to analytic calculations;
- Peak attenuation: the ratio of the maximum of the direct image (at $\rho = 0$) to that of the coronagraphic image;
- Contrast: the ratio of the maximum of the direct image to the intensity of the coronagraphic image at a given separation ρ, azimuthally averaged.

These metrics are not necessarily related, so it is important to consider all of them to accurately assess the performance of the coronagraph.

Fig. 4 (Riaud et al. [6]) shows a comparison of the FQ-PM nulling performances with and without a residual atmospheric tip/tilt of about 10 mas, which must be considered as small in ground-based astronomy. As soon as a residual tip/tilt occurs, even a fairly small one, the nulling performance is strongly degraded. In our example, the tip/tilt of $0.088\lambda/D$, i.e. 8.4 mas on a 3.6 m telescope at 1.65 μm, leads to a degradation of typically 12 magnitudes. The residual tip/tilt gives clearly the major contribution in the loss of nulling.

We have simulated an even more realistic case by introducing in the model a corrugated wavefront representative of that found after correction by an adaptive optics system on a ground-based telescope. This is illustrated on Fig. 4-right, where the situation we have studied corresponds to a median Strehl ratio of 0.74, variable seeing and a central obscuration of 25% in diameter. An 8 magnitudes fainter companion is distinguishable after co-addition of two thousand frames. This range of 8 to 12 magnitudes at $3\lambda/D$ for the star/companion contrast must be considered as the typical performance that one can aim at from the ground. From space, it is more hazardous to predict a precise

performance, since the quality of the optics will certainly be the limit on the phase error. It is extremely unlikely that the maximum theoretical gain be obtained easily, because it would correspond to phase errors in the μ-radian range; however, a significant step towards this goal could be achieved.

3.2. Image processing

Image processing is another way to improve the rejection rate of the spurious stellar light. One notices that, as long as the phase variance is not too large (i.e. $\exp(i\phi) \sim 1 + i\phi$), the speckle image produced with the FQ-PM is centro-symmetric: this is due to the mask symmetry and its attenuation capability. Such an assumption can be shown by considering the properties of the Fourier transform regarding the parity (Boccaletti et al. [7]). We can take advantage of this property when looking for the detection of a faint companion, since it should appear as an asymmetric pattern in the image. If we compute the centro-symmetric image and subtract it from the original image we then strongly reinforce the contrast of the single feature. This technique was applied in the example given in Fig. 4.

3.3. Manufacturing of the 4QPMC

For a basic device, the π phase shift is obtained with a simple step: the difference in thickness (e) for two adjacent quadrants is such that $(n-1)e = \lambda/2$. An estimate of the required accuracy on the thickness can be obtained by considering that the residuals due to the device itself must be below the error due to residual random tip/tilt (Riaud et al. [8]). The phase difference induced by the device (error of thickness or index variation) produces a residual amplitude equal to the phase error angle times half the amplitude of the incoming wave. In other words, the relative error on the amplitude is equal to half the phase error expressed in radians. For instance, if the AO system and the seeing allow a rejection factor of 10 magnitudes (thus 10^{-2} on the amplitude) with a perfect mask, then, for the actual mask, the accuracy on the phase difference between two quadrants must be of 2×10^{-2} radian, corresponding to a few nanometers in thickness: this does not appear so stringent. Note that, to minimise the effect of the surface errors at low frequencies (for instance, gap between quadrants), the Airy pattern can be enlarged (the question of the field of view is not actual: in general a few arc-second are sufficient), so that a large magnification is used to produce the first image on the FQ-PM.

As regards the question of manufacturing, three techniques have been mainly considered: lift-off, ion etching and chemical vapour deposition.

- *The lift-off process:* The mask is simply a transmission plate where a transparent coating is deposited. The specific quadrants geometry is obtained by the *lift-off* technique, a process with sequential steps: (i) deposition of a photosensitive resin layer on the substrate; (ii) ultraviolet illumination through a photolithographic mask with two opaque quadrants; (iii) resin development and removal of exposed resin; (iv) deposition of SiO_2 with the Ion Assisted Deposition technique (Fresnel Institute in Marseilles); (v) chemical removal of the unexposed resin layer plus its SiO_2 overcoating. A quadrant transition thinner than 0.5 μm can be obtained. Instead of manufacturing a single FQPM, it may be wise to have a substrate supporting an array with many individual masks and to choose the one giving the best results. In addition, a thin chromium cross (10 μm in width) was deposited on the masks we used, to avoid some nulling degradation induced by the imperfect quadrant transition.
- *Ion etching:* In this technique, no additional material is deposited to make the step: on the contrary, there is an attack of the substrate which is done using a ionic gun. The most efficient process appears to be association of energetic ion bombardment with a reactive gas. The control of the gun that gives the accuracy of the engraving. Germanium, for instance, can be used efficiently with this process (the case of the JWST/MIRI coronagraphs done at CEA).
- *Chemical vapour deposition:* This technique is especially well adapted to diamond and gives good results when associated to O_2 ion etching technique (several samples manufactured by Uppsala University).

3.4. Laboratory results

Several specific benches, one in the visible and a cryogenic one in the infrared, were developed to demonstrate the 4QPM effect and assess the performance.

The visible experiment demonstrated rapidly (Riaud et al. [8]) that a deep extinction of the stellar peak (44 000) and a very low residual speckle level (10^7 to 10^6) can be reached with standard-quality lenses. The limitations are various, but the wavelength shift, together with the surface quality, seems to be the most important cause of degradation factors. The spectral bandwidth contributes much less to the degradation of the rejection factor. Since then, some improvement have permitted to reach a better rejection factor of 1.2×10^5 for the peak and a contrast of 10^6 at $3\lambda/D$.

With the infrared bench, in a broadband ($R = 9.5$), we recorded the following contrast with a device optimised for 4.8 μm: 1.12×10^3, 5.1×10^4 and 3.3×10^5 at respectively 1, 3 and 6 λ/D. Note that those figures are extremely close to the predictions by a numerical model that took into account the various parameters of the test. This gives a fair level of confidence for the future.

3.5. Variants of the 4QPM

We briefly review several variations on the concept of 4QPMC that have been proposed:

- *The hybrid Lyot/4QPMC:* We have first envisaged the use of a hybrid mask that will change both the phase with the four-quadrant mask and the amplitude with a small Lyot mask in the centre. Results of our simulations made with this hybrid solution show a significant reduction of the induced tip/tilt perturbation. In addition, the 4QPMC we developed and used were manufactured with a deposited thin opaque line at the frontier between quadrant, in order to mask defaults (small scratches) along the edges of the quadrants.
- 8, 16, 2^N *quadrant phase mask coronagraphs:* The idea is simply to multiply the number of sectors, with the rule that it must be a power of two: thus 8-, 16-sectors phase mask coronagraphs can be imagined and should be in fact very efficient devices. The nulling efficiency improves directly as a power of the number of sectors divided by two. This can be understood by considering for instance the 8-sector phase mask coronagraph as two 4QPMC, one following the other, with a rotation of $\pi/4$. The destructive interference effect is applied twice. Numerical simulations confirm this behaviour. Of course, there is a drawback which is the multiplication in the image plane of the frontiers between sectors and thus a loss of space available for the searched circumstellar features, such as planets for instance. However, the case of the 8-quadrant deserves certainly more attention than it has received up to now: it may represent the best compromise between a efficient nulling and a still acceptable free zone.
- *The phase knife coronagraph:* In this solution (Abe et al. [9]), as in the achromatic design using stack of plates (see below), the phase mask is an assembly of phase plates of equal thickness but different optical refraction indices. However, instead of using four separate plates, they use a combination of two crossed pairs, each one called a phase knife (Abe et al. [10]). The combination of indices and glass thickness are optimised to obtain a theoretical nulling effect of $\sim 10^2$ over a spectral band ranging from $\lambda = 450$ to 620 nm. The two orthogonal pairs of phase knives are held together between two identical glass blocks. This optical design solution was chosen to circumvent the difficult operation of shaping and assembling four independent quadrants. However, on the first prototype, some difficulties were found due to the molecular bound and also because the differential thickness between the two plates should be less than a few nanometers, which is a stringent condition.
- *The elliptical 4QPMC:* One possible solution to overcome the problem of the central obstruction which reduces notably the performances of the 4QPMC, could be to consider only a smaller but filled sub-pupil on the primary mirror of the telescope. Haguenauer et al. [11] presented preliminary results of this technique. Of course, the price to pay is a reduction of the pupil diameter and thus of the angular resolution. However, a compromised solution can be envisaged by defining an elliptical pupil that can fit the largest possible filled area on the primary. In that case, the 4QPMC would have to be designed accordingly: it would have quadrants that are no longer at right angles and a dedicated orientation with respect to the elliptical shape of the pupil, but the principle is the same and the high rejection efficiency will be kept. Of course the drawback is that the angular resolution cannot be the same in all directions.

4. Instrumental developments and performance at telescope

Today, FQPMC have been installed by our group on one operational instrument, NACO, the adaptive optics system of the VLT, and will be integrated in two instruments currently in development: the mid-infrared camera MIRI—one the three instruments of the JWST—and SPHERE, the future planet finder instrument of the VLT.

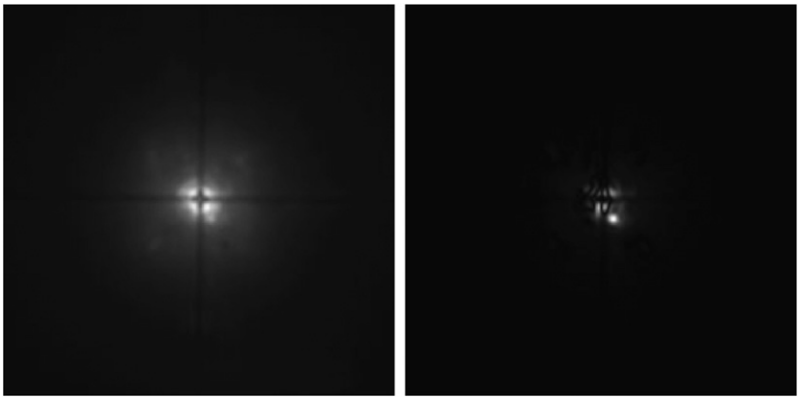

Fig. 5. Coronagraphic images (raw and subtracted) of AB Dor A obtained with the new 4QPMC installed on NACO and used in conjunction with the SDI sub-system. The companion AB Dor C is clearly visible. The display is not linear ($I^{0.5}$) on the raw image (left), as opposed to the subtracted image (right).

4.1. VLT-NACO

The first high-contrast images, using a four-quadrant phase mask coronagraph, have been obtained at the ESO Very Large Telescope. In agreement with ESO, we installed in August 2003 an infrared FQPM inside the near-IR camera CONICA of the NACO adaptive optics imaging system of the VLT. The FQPM is optimised for the K band where correction is expected to be the best. Given the seeing conditions, achromaticity is not required here, since the nulling degradation introduced by the residual phase defects after adaptive optics compensation is about 30 times larger than the limitation set by the chromatism. For the same reason, the presence of a central obscuration in the telescope aperture is not critical.

The FQPM was manufactured by the Fresnel Institute. A layer of SiO_2 was deposited onto a SiO_2 substrate. The four quadrant design was obtained with a photolithographic technique, achieving a lateral accuracy of a few microns. The two-night commissioning was carried out in January 2004 (Boccaletti et al. [12]). We evaluated the behaviour of the coronagraph on a variety of astrophysical targets: binary stars, circumstellar disks, and active galactic nuclei. The performance of the coronagraph is in agreement with our expectations based on numerical simulations. Despite moderate correction, the phase mask provided a stellar peak attenuation of a factor of about 10 on average (for long exposure), and its performance is limited only by the phase residuals (mainly low-order aberrations) that are left uncorrected by the adaptive optics system.

A second 4QPMC was recently installed in NACO to provide a better efficiency when using the SDI system (Stellar Differential Imaging) system to perform high contrast imaging. This 4QPMC, optimised for the H-band, was manufactured by GEPI through reactive ion etching technique. Fig. 5 shows the results obtained with this last device on the double star AB Dor: the (known) low mass companion C is clearly visible.

Note that several astrophysical results have already been obtained with this instrumental mode of NACO: e.g. the first worldwide use of a coronagraph in extra-galactic work, with the detection of structured emission in the NAG NGC1068 (Gratadour et al. [13]), or the detection of a debris disk around a T Tauri star (Riaud et al. [14]).

4.2. SPHERE

For its second-generation instrumentation on the VLT, ESO has supported two phase A studies for a so-called *Planet Finder* dedicated instrument. Based on the results of these two studies, a unique instrument, SPHERE, is now considered for first light in 2010, including a powerful extreme adaptive optics system, various coronagraphs, an infrared differential imaging camera, an infrared integral field spectrograph and a visible differential polarimeter. This instrument is developed by a large European consortium with the Observatoire de Grenoble at its head. Our team in Meudon is responsible for the coronagraphic sub-system and will provide specifically achromatic 4QPMC.

The baseline coronagraph suite will include an achromatic four-quadrant phase mask coronagraph (A4QPM) based on precision mounting of four half-wave plates (see below), and both a classical Lyot coronagraph and an apodised

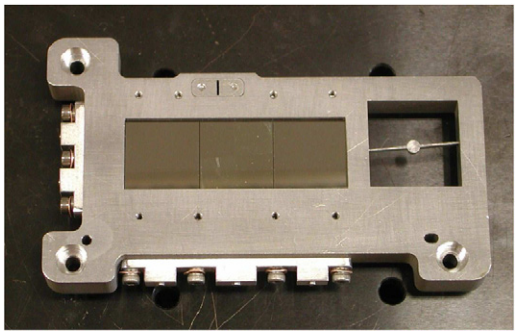

Fig. 6. The set of 3 4QPMC and one Lyot coronagraphs that have been manufactured and delivered to be mounted in the MIRI/JWST camera.

Lyot coronagraph. The A4QPM has recently been demonstrated in the visible, where the main difficulties of precision edge polishing and mounting of the half-wave plates have been addressed and excellent performance has been demonstrated (Mawet et al. [15]). An alternative, also using the principle of half wave plates, is the use of sub-wavelength engraved devices also called Zero Order Gratings or ZOG (see below): this is extremely promising because it is a monolithic component with no problems of assembly; however, given the fact that the demonstration is not yet fully complete, this solution is not considered as the baseline.

4.3. MIRI-JWST

The James Webb Space Telescope (formerly the Next Generation Space Telescope) will be capable of unprecedented science, owing to its large diameter, its low background, and its dedicated IR instruments. One of these instruments is a mid-IR facility called MIRI (Mid-IR Instrument), which is being developed through a collaboration of US and European teams. Our group at the Observatoire de Meudon has developed the coronagraphic device for the MIRI camera, Boccaletti et al. [16]. Taking advantage of the reduced star-planet contrast in this spectral range, the prime goal of this coronagraph is the direct detection of extrasolar planets. Differential imaging in several well-chosen spectral windows is foreseen.

MIRI is made up of two modules, a camera where the coronagraph is installed and an integral field unit spectrograph (Wright et al. [17]). The coronagraph masks are located at the JWST focal plane (the entrance of the MIRI camera). The Lyot stops are set in the filter wheel, and each is associated with a single filter and a single coronagraph. There are actually four coronagraphs inside the MIRI camera: one standard Lyot mask operating at 23 μm, optimised for cold objects, such as circumstellar disks, plus three monochromatic 4QPMC at 10.65, 11.40, and 15.50 μm, respectively, optimised for the detection and characterisation of Extrasolar Giant Planets. These filter wavelengths allow us to derive some physical parameters of the EGPs, such as the temperature of the planet and the abundance of ammonia in the atmosphere. This information is essential for comparing evolutionary models with actual data.

The 4QPMs that are chosen for MIRI are monochromatic. A filter resolution of $R = 20$ has been chosen so that the chromatic attenuation reaches 2000 with a standard 4QPM: this is better than what is expected, because telescope defects will dominate the error budget. The performance is however compatible with the measurement of ammonia absorption in the EGP atmosphere.

In order to choose the optimal FQPM for the MIRI camera, several FQPMs were manufactured with different processes, using the most favourable materials in the mid-IR: germanium (Ge), zinc selenide (ZnSe), and chemical vapour deposition diamond. After a thorough phase of testing in cryogenic conditions with a dedicated test bench, the solution of germanium substrate on which engraving of the quadrants is done with reactive ion etching, was retained.

We have demonstrated (Baudoz et al. [18]) that the behaviour of the four-quadrant phase mask in the mid-IR and at a very low temperature (12 K) corresponds to the expected performance. The fully integrated system (see Fig. 6) has been delivered and installed in the MIRI camera. It includes two different sets: an assembly with the three 4QPMC and one Lyot mask at the focal plane and a set of dedicated pupil masks of very peculiar shape—in order to comply with the segmented telescope of JWST—mounted on the filter wheel.

5. The achromatization issue

The question of achromaticity is an important one, since if we consider the most demanding goal (exoplanet detection), the flux of the faint companion to be detected is extremely low, which requires the use of the broadest wavelength band. The basic device designed with a simple step, is intrinsically monochromatic. When a finite wavelength $\Delta\lambda$ is used, one can easily show, using a basic Fresnel's vector representation, that the efficiency of the 4QPMC—defined as the ratio of the total integrated intensity with and without the FQPM—is $\tau = (48/\pi^2)R^2$, where $R = \lambda/\Delta\lambda$. This means that with a classical broad-band filter, of typically $R = 5$, the expected efficiency is only 125. A serious issue is thus to obtain an achromatic π phase-shift device on an extended wavelength range. Note that this is the same problem met in nulling interferometry.

Several solutions have been investigated in order to solve the problem: they are briefly described now.

5.1. Stack of plates

Riaud et al. have examined the possibility of manufacturing stacks of plates of different materials and different thickness in order to compensate the harmonic dependence versus lambda of the phase, by the linearization of the dispersion law in the expression: $\phi(\lambda) = \sum_k (n_k - 1)e_k/\lambda$. In fact two plates are sufficient in all cases to reach a fairly good achromatization and Fig. 7 illustrates the design with only two materials.

In the thermal infrared some very achromatic solutions do exist for a transmitting masks (as well as for a reflection mask: see Riaud et al. [6]). For example, a solution with dispersive elements, using the couple AgCl and ZnSe—well suited for a 10 μm broad-band filter—, provides variance of the phase of 4.4×10^{-3} rms. Other combinations are possible, for instance, MgF_2/BaF_2. The overall nulling factor would be about 50 000 for a perfect off-axis telescope, allowing in principle the detection of a planet with a contrast of 10^{-6} at an angular distance of a few λ/D.

There are two basic difficulties with this solution in terms of manufacturing: the first is that a stack of plates of different materials means than we have to manufacture separate plates with extremely precise edges that must be put side by side. The second is that the manufacturing of these stacks of plates requires an accuracy in thickness which is extremely high, since it is no longer a relative difference of thickness which is looked for, as in the monochromatic case, but an absolute value of the thickness for each plate (all are of different thickness), with an accuracy of a few nanometers: this means extremely high constraints on the knowledge of the material index, especially if one aims to work at cryogenic temperature, and a process of manufacturing as well as a supporting design extremely demanding. Abe et al. [19], who report the design and use of a system of this type, confirm the difficulty of manufacturing such a device.

5.2. Multi-layer coating

The use of photolithography should allow one to deposit layers of different thickness on different areas of the substrate. With a technique similar to that developed for anti-reflection coatings, obtaining a precise thickness should not be a problem.

A multi-layer coating (see Fig. 7-right) with a proper combination of material and thickness, again as made for a broad-band anti-reflection coating, should produce an achromatic mask with a phase error small enough to be below the residual of the corrected wavefront of an AO system. The condition that must be verified by the mask

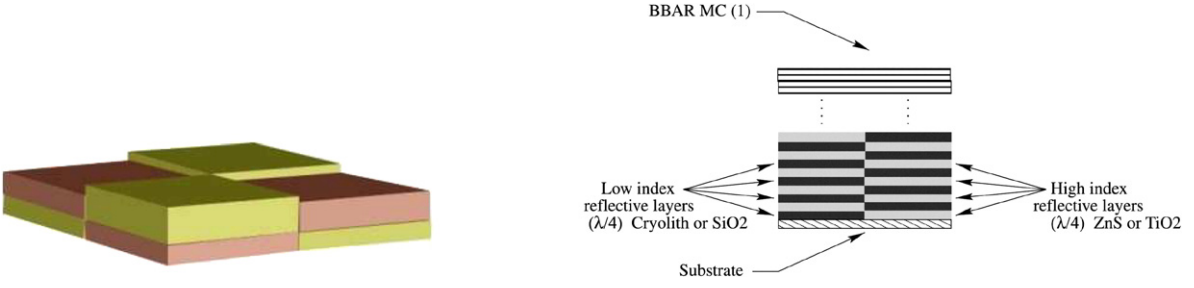

Fig. 7. Right: Principle of the achromatic 4QC using stacks of two materials. Left: Principle of achromatization using multi-layer coatings.

with its different layers is that the quadratic deviation from π of the phase difference be minimum in the considered wavelength range. If a is the index for the π phase-shift quadrants and b for the other ones, and e_i and $n_i(\lambda)$ the thickness and refraction index of layer i, then this condition can be expressed by minimising the expression:

$$\int_{\lambda_1}^{\lambda_2} \sum_{i=1}^{k} \left[\phi_i^a(\lambda) - \phi_i^b(\lambda) - \pi\right]^2 \mathrm{d}\lambda = \int_{\lambda_1}^{\lambda_2} \sum_{i=1}^{k} \left[\frac{2\pi}{\lambda}\left(\left(n_i^a(\lambda)-1\right)e_i^a - \left(n_i^b(\lambda)-1\right)e_i^b\right) - \pi\right]^2 \mathrm{d}\lambda$$

5.3. *The* 0, 1, 1, 2 *four quadrant design*

In this design, the FQPM is slightly different: three of the quadrants are as in the original design, i.e. two quadrants with a phase shift of π and one with no phase shift. The difference is on the fourth quadrant which applies a 2π phase shift instead of no phase shift. Why is such a design better in terms of achromaticity? The answer is found when one looks at the Taylor's development of the amplitude after the light has propagated through the phase mask. This amplitude reads:

$$a = 1 + 2 \times \exp j\phi + \exp j2\phi$$

where ϕ is the phase shift introduced by each step at a wavelength which is no longer λ_0: $\phi = 2\pi(n(\lambda) - 1)e/\lambda$, where e is the thickness of one step and $n(\lambda)$ is the index. Of course for $\lambda = \lambda_0$, we have: $\lambda_0/2 = (n(\lambda) - 1)e$, so that $\phi = \pi\lambda_0/\lambda$, which is close to π when λ is close to λ_0: $\phi = \pi(1 + \delta\lambda/\lambda)$. Noting $\varepsilon = \pi\delta\lambda/\lambda$, we can thus make the Taylor's development of a:

$$a = 1 - 2\left(1 + j\varepsilon - \varepsilon^2 + \cdots\right) + \left(1 + j2\varepsilon - 4\varepsilon^2 + \cdots\right)$$

The terms of degree 1 in ε cancel out and the dependence versus λ is in ε^2 thus largely improved with respect to a classical design. One drawback of this solution, which is currently under study for manufacturing a device, is that a tilt is introduced on the image, with a certain asymmetry of the residual.

Note that the improvement can be even better when considering devices with more that 4 quadrants (this design has already been evoked in the previous section). For instance a component designed with 8 sectors whose steps are respectively [1, 2, 1, 2, 1, 0, 3, 2], will have a rejection varying as ϕ^3: as an example, with $\Delta\lambda/\lambda = 0.1$, the rejection factor is still of magnitude 3×10^{-4}.

5.4. *The multi-stage 4QPMC*

Here it is proposed to use a combination of chromatic FQPM in cascade to achromatize the overall dephasing of the coronagraph by properly choosing the central wavelength of each component. This way, a rather broad band can be split in sub-band, each being controlled by one 4QPMC. A preliminary theoretical study of the achromatization using this Multiple-Stage FQPM shows that very a promising performance can be obtained. For instance, in the whole H-band (1.65 μm, $R = 4.7$), a nulling of 14 000 is reached in perfect conditions. A test bench is currently designed and assembled in Meudon in order to test this new concept.

5.5. *Half-wave plates*

An elegant solution of the question of achromaticity is to use for the quadrants achromatic half-wave plates, where two of them are rotated of 90°. The principle is illustrated on Fig. 8.

The achromatization of the FQPM by achromatic half-wave plates (HWP) can be explained as follows. Let s and p be the global polarisation components of the incoming light. In each of the four quadrants made of cut half-wave plates, the s and p global polarisation states are projected according to the local fast and slow axis orientations of the two-level stack. Two indices, the ordinary index n_o and the extraordinary index n_e, can be assigned to these directions. Let us now assume that the four cut quadrants are strictly identical (same two-level stack) and that two of them, opposite along one diagonal, are rotated by 90° with respect to the two others. This anti-symmetrical configuration mimics the FQPM particular focal plane π-phase distribution for each parallel and potentially interfering state of polarisation (see Fig. 8). Such a design works with natural light. In practice, all plates have to be polished and cut in the same material blank to obtain homogenous components in terms of refractive index and surface quality.

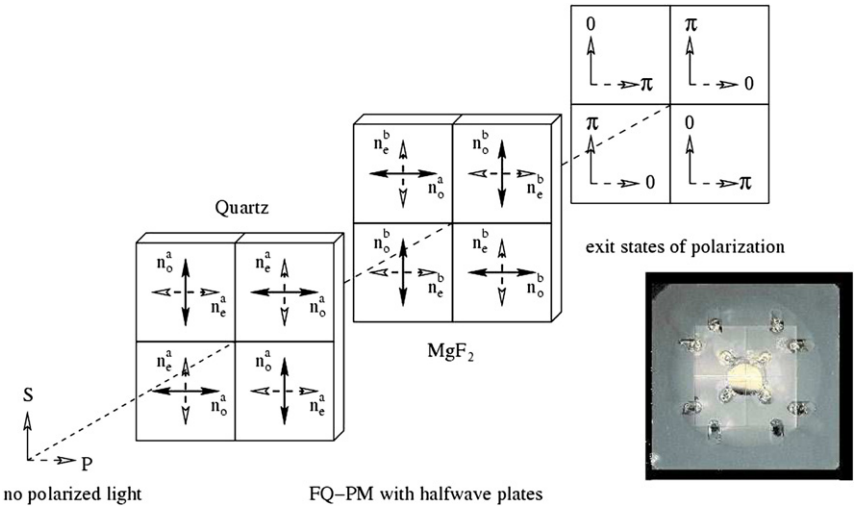

Fig. 8. Principle of the achromatic 4QC using half wave plates.

These phase shifters combine birefringent plates made of different materials with properly chosen thickness. The HWP thickness control is relaxed by two orders of magnitudes with respect to the classical (non-birefringent) dispersive plate approach, a huge advantage!

In the experimental demonstrator developed at Meudon (Mawet et al. [15]), we used a two stage stack of Quartz and MgF_2. This combination allows to cover a large spectral range in the visible (500–900 nm) with a small phase error residual around π (≈ 0.12 rad rms). With this achromatization, we obtained an attenuation of 755 on the white light PSF peak. This solution is directly applicable to ground-based telescopes using high order adaptive optics: it is indeed the baseline of SPHERE, the ESO's VLT *planet finder* project and could easily be transposed in the mid-infrared domain for future space-based missions like DARWIN/TPF.

Another way to make the FQPM achromatic was proposed and tested by Murakami et al. [20]: the polarisation interferometric coronagraph. It is in fact a FQPM whose phase shift is provided by a Liquid-Crystal (LC) device sandwiched between two crossed polarisers. This method allows a broadband use of the coronagraph, but only with linearly polarised light in the visible.

5.6. Zero-order gratings: the 4QZOG and the AGPM

The design that is today probably the most promising was introduced by one of us (D. Mawet): the idea is basically the same as the one used in half-wave plates, but here the difficulty of assembling pairs of plates at the micron scales is solved by engraving the birefringent property in the substrate of a unique plate, Mawet et al. [21]. When the period of a grating is smaller than the wavelength of the incident light, it does not diffract as a classical spectroscopic grating. All the incident energy is enforced to propagate only in the zeroth order, leaving incident wavefronts free from any further aberrations. The sub-wavelength gratings are therefore often called zeroth order gratings (ZOGs). This type of gratings behaves like homogeneous media with unique characteristics, which can be used to synthesise artificial birefringent achromatic wave-plates.

Intuitively, one can understand this artificial anisotropy and the existence of two distinct effective indices: the incident light sees two different media as its vectorial components vibrate parallel or orthogonal to the grating lines. The idea to obtain an achromatic half-wave plate, is to make the form birefringence proportional to the wavelength in order to achromatize the subsequent differential phase shift between the two polarisation states TE and TM.

In the case of the FQPM implementation, the four gratings engraved on a unique substrate are strictly identical and implemented in the following way: two of them in two quadrants along one diagonal are rotated by 90 degrees around their normal with respect to the two others. This anti-symmetrical configuration achieves the FQPM particular focal plane phase distribution (Fig. 9, left). This new component is referred as to the Four Quadrant-ZOG, or 4QZOG (Mawet et al. [5]).

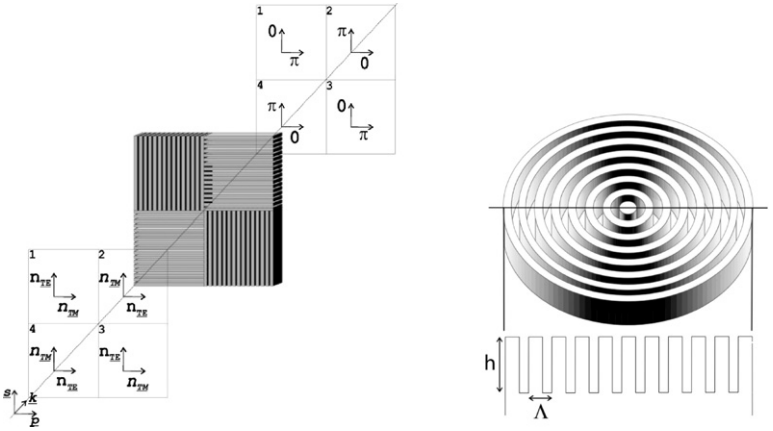

Fig. 9. Principle and structure of the 4QZOG (left) and of the AGPM (right).

The annular groove phase mask (AGPM) is a very clever variant of the 4QZOG: a space variant ZOG (Fig. 9, right) synthesises a spiral phase plate which creates an optical vortex, Mawet et al. [5]. Indeed, at the centre of the components, the phase possesses a screw dislocation inducing a phase singularity, i.e. an optical vortex. The central singularity forces the intensity to vanish by a total destructive interference, creating a dark core. This dark core propagates and is conserved along the optical axis. It can be demonstrated that a focal plane vortex leads to a total rejection in the ideal case (Mawet et al. [22]). Furthermore, the unique ZOG properties permit a broadband use, as in the case of 4QZOG. Indeed, the ZOG parameters (period, depth, filling factor) are the same in the 4QZOG and AGPM configurations. The only difference concerns the geometry of implementation which suppress the dead zones.

For example, the coronagraphic profile in the K band shows a peak-to-peak null depth better than about 10^{-5}. The speckle level of 10^{-7} is quickly reached at a few λ/D. The AGPM coronagraphic behaviour is very similar to the achromatic 4QZOG coronagraph, but with a total symmetry. As mentioned before, the 4QPM/4QZOG quadrant transitions induce a non-negligible attenuation of the circumstellar features lying on them. These dead zones represent 10% of the focal plane (at $6\lambda/D$). Thanks to the perfect AGPM circular symmetry, this problem no longer exists. Note that these performances are comfortably one order of magnitude above the H-band/K-Band VLT-SPHERE specification, for example.

6. Conclusions

Direct detection and spectral characterisation of extrasolar planets is one of the most exciting but also one of the most challenging areas in modern astronomy. High contrast imaging and especially coronagraphy is one of the key technology to master in order to obtain direct images of those planets. The introduction a few years ago of the new concept of the Four Quadrant Phase Mask Coronagraph has significantly contributed to the advance in that field. Its advantages are simplicity and potentially an extremely efficient rejection capability. Curiously for such a simple system, it has motivated many new ideas, several of them being currently in the process of development and testing. Will one of those new concepts be part of the lucky experiment that will make the first image of a planet around a main sequence star? Nobody knows ...

References

[1] D. Rouan, in: S. Udry (Ed.), Planetary Systems and Planets in Systems, ESA Publications Division, 2007.
[2] P. Baudoz, Y. Rabbia, J. Gay, Astron. Astrophys. Suppl. 141 (2000) 319.
[3] F. Roddier, C. Roddier, Publ. Astron. Soc. Pacific 109 (1997) 815.
[4] D. Rouan, P. Riaud, A. Boccaletti, Y. Clénet, A. Labeyrie, Publ. Astron. Soc. Pacific 112 (2000) 1479.
[5] D. Mawet, P. Riaud, O. Absil, J. Surdej, Astrophys. J. 633 (2005) 1191.
[6] P. Riaud, A. Boccaletti, D. Rouan, F. Lemarquis, A. Labeyrie, Publ. Astron. Soc. Pacific 113 (2001) 1145.
[7] A. Boccaletti, P. Riaud, D. Rouan, Publ. Astron. Soc. Pacific 114 (2002) 132.
[8] P. Riaud, A. Boccaletti, J. Baudrand, D. Rouan, Publ. Astron. Soc. Pacific 115 (2003) 712.

[9] L. Abe, A. Domiciano de Souza Jr., F. Vakili, J. Gay, Astron. Astrophys. 400 (2003) 385.
[10] L. Abe, F. Vakili, A. Boccaletti, Astron. Astrophys. 374 (3) (2001) 1161–1168.
[11] P. Haguenauer, E. Serabyn, B. Mennesson, et al., in: Space Telescopes and Instrumentation I: Optical, Infrared, and Millimeter, in: Proceedings of the SPIE, vol. 6265, 2006, p. 62651G.
[12] A. Boccaletti, P. Riaud, P. Baudoz, et al., Publ. Astron. Soc. Pacific 116 (2004) 1061.
[13] D. Gratadour, D. Rouan, A. Boccaletti, P. Riaud, Y. Clénet, Astron. Astrophys. 429 (2005) 433.
[14] P. Riaud, D. Mawet, O. Absil, et al., Astron. Astrophys. 458 (2006) 317.
[15] D. Mawet, P. Riaud, J. Baudrand, et al., Astron. Astrophys. 448 (2006) 801.
[16] A. Boccaletti, P. Baudoz, J. Baudrand, J.M. Reess, D. Rouan, Adv. Space Res. 36 (2005) 1099.
[17] G.S. Wright, F. Bortoletto, C.F. Bruce Jr., et al., in: J.C. Mather (Ed.), IR Space Telescopes and Instruments, in: Proceedings of the SPIE, vol. 4850, 2003, pp. 493–503.
[18] P. Baudoz, A. Boccaletti, P. Riaud, et al., Publ. Astron. Soc. Pacific 118 (2006) 765.
[19] L. Abe, M. Beaulieu, F. Vakili, et al., Astron. Astrophys. 461 (2007) 365.
[20] N. Murakami, N. Baba, T. Ishigaki, in: S. Ikeuchi, J. Hearnshaw, T. Hanawa (Eds.), 8th Asian-Pacific Regional Meeting, vol. II, 2002, pp. 55–56.
[21] D. Mawet, C. Lenaerts, V. Moreau, et al., in: A.B. Schultz (Ed.), High-Contrast Imaging for Exo-Planet Detection, in: Proceedings of the SPIE, vol. 4860, 2003, pp. 182–191.
[22] D. Mawet, P. Riaud, O. Absil, J. Baudrand, J. Surde, SPIE (2005) 59051N, 1–10.