Focal spot generator without sidelobes and its application in coronagraphy

GORDON D. LOVE† and DAVID F. BUSCHER‡
†University of Durham, Rochester Building, Dept. of Physics, Durham, DH1 3LE, UK; email: g.d.love@durham.ac.uk
‡University of Cambridge, Cavendish Laboratory, Madingley Road, Cambridge, CB3 0HE, UK

(Received 28 January 2004)

Abstract. A combination of two phase filters is described, which together produce a light beam that can be focused to produce a focal spot with an approximately flat top and with much reduced sidelobes (a top-hat function). There are a variety of potential applications, but our interest lies in the use of this device in astronomical coronagraphy and point spread function engineering in order to aid the detection of extra-solar planets.

1. Introduction

There have been a large number of papers written on the efficient production of flat-topped light beams with minimal sidelobes. The achievement of such beams is important for areas ranging from laser machining and lithography to the production of light tweezers and optical traps. There are a variety of methods of producing flat-topped beams, for example, diffraction [1–3], adaptive optics [4, 5], refraction [6], or simply defocusing the beam [7]. There has also been interest in the related topic of super-Gaussian beams from a more theoretical perspective [8–10]. These papers describe methods for producing flat-topped beams in both the near and far field.

The application of interest here is in the direct imaging of planets around stars other than the Sun (extra-solar planets). The main challenge here is the detection of a faint companion which can be up to nine orders of magnitude fainter than the star. Even if the effects of atmospheric turbulence and static aberrations have been completely removed by adaptive optics, or the telescope is in space, then diffracted light in the wings of the star’s image will completely obscure the planet’s image. A whole variety of techniques have been proposed to cancel the light from the star including nulling interferometry and a whole range of coronagraphic techniques (see [11]). Engineering the point spread function so as to produce very low sidelobes has also been proposed using, e.g., square telescopes and apodization [12, 13]. Here we propose a technique which produces an image of the star with very low sidelobes, so as to enhance the planet detection. An assessment of its use in coronagraphy is outside the scope of this paper. However, it is important to note that in this application the input distribution of light is a top-hat function (the telescope pupil) whereas in many of the earlier references on flat-topped beam

*Author for correspondence.
production the input beam was assumed to be Gaussian. Thus, the maintenance of a top-hat function between the pupil and image planes of a system is useful (although our system only does this for one pair of conjugate points).

Our proposed method involves producing a focal spot from the star in the far field from a uniformly illuminated aperture using two phase masks which can produce an approximately top-hat focus which has very low subsidiary maxima and a quasi-flat-top. (In coronagraphy, this light can then easily be blocked with a mask.) The phase masks do not absorb any of the energy in the beam and so this method is both simple and efficient.

2. Methodology

This design was influenced by the phase-shifting coronagraph described by Roddier and Roddier [13], which in turn is rather similar to a Zernike phase contrast experiment (in its set up, rather than its purpose).

The top-hat generator is based on the following conceptual idea. The Airy function, with its Gaussian-like central peak and subsidiary maxima, is produced by focusing a uniform cross-section light beam. This process can be described mathematically by the square modulus of the Fourier transform of the top-hat aperture function. Therefore, in order to create a top-hat focal spot, a mechanism to create a pupil function with an Airy-like (amplitude) structure is needed. This will then be transformed when focused to produce a top-hat focal spot. An Airy-like pupil function (with the correct amplitude and phase) could, in principle, be produced by a combination of amplitude and phase masks, or one or more spatial light modulators. However, this would be both complicated and wasteful of light. An alternative method is to use two phase masks as described here.

The conceptual optical arrangement is shown in figure 1. The input light beam is focused by the first lens to produce the usual Airy disk. This is incident on the first phase mask, which consists of a circular phase-shifting spot of radius $s$, where $s$ is smaller than the radius of the central peak of the Airy spot. The light then propagates to a second lens where it is recollimated and where the second phase mask is placed. The size of the phase shifting spot on this second mask is the same size as the image of the first aperture. Note that the sizes of the second two lenses must be much larger than the diameter of the reimaged aperture. A third lens then re-focuses the beam and produces the final top-hat focal spot.

The operation of the system can be understood qualitatively by means of a simplified 1-dimensional analysis shown schematically in figure 2. The top-hat input aperture function shown in figure 2(a) is Fourier transformed to produce the ‘sinc’ amplitude distribution in the plane just before the first phase mask, shown in figure 2(b). The field just after the phase mask (shown in figure 2(c)) can be considered as the difference of two components (according to Babinet’s principle): the first is the field that would be present if the phase mask was not there, and the second is twice the field that would be present if the whole mask were opaque apart from the area corresponding to the phase spot. The factor of two arises because the mask is a phase mask and not an amplitude mask.

The field distribution just before the second phase mask is a Fourier transform of the field after the first phase mask. The linearity of the Fourier transform means that the two components of the latter field can be separately transformed. The first component transforms to give a scaled version of the original pupil function.
The second component can be approximated as a top-hat function, as long as the size of the phase spot is small compared to the width of the sinc function. Therefore when it is transformed it produces a sinc function that is much broader than the pupil function. The difference of these two terms is shown in figure 2(d).

It can be seen that this is a sinc function with a central discontinuity whose width is proportional to that of the original pupil function. The second phase mask flips the sign of this discontinuity relative to the rest of the function. If the amplitude of the two field components (after the first phase mask) is carefully optimized (by adjusting the size of the phase spot) then the absolute value of the field in the centre of the pupil will be equal to that from the sinc function. After the passage through the second phase mask, the field is then flipped to produce a quasi-sinc function, as shown in figure 2(e). This is then transformed by the final lens to produce the top-hat focal spot, shown in figure 2(f).

In the next section we perform a mathematical analysis in order to calculate the size of the required phase spot.

3. Analysis

The incident pupil function, \( A_0(x, y) \), is described by a circular top-hat function of diameter \( d \) and unit amplitude

\[
A_0(x, y) = \text{circ}\left(\frac{2d}{d}\right),
\]
where \( x \) and \( y \) are the Cartesian coordinates in the pupil plane, \( r = \sqrt{x^2 + y^2} \) is the radial coordinate and the circ function is described by,

\[
\text{circ}(r) = \begin{cases} 
1 & r < 1 \\
0 & r > 1 
\end{cases}
\] (2)

The distribution of light in the first focal plane, \( A_1(u, v) \) (the plane of the first phase mask) is proportional to the Fourier transform of \( A_0(x, y) \) (see, for example, [12]), and is given by

\[
A_1(u, v) = \frac{1}{i\lambda f_1} \int_{-\infty}^{\infty} A_0(x, y) \exp \left( \frac{i2\pi}{\lambda f_1} (xu + yv) \right) dx dy,
\] (3)

where \( f_1 \) is the focal length of the first lens, \( \lambda \) is the wavelength, and \( (u, v) \) are the Cartesian coordinates in the focal plane. A quadratic phase factor has been omitted, as this does not affect this calculation. The light distribution in the first focal plane is therefore given by

\[
A_1(u, v) = \frac{1}{i\lambda f_1} \frac{\pi d^2}{4} \text{jinc} \left( \frac{\pi d \rho}{\lambda f_1} \right),
\] (4)

where \( \rho = \sqrt{u^2 + v^2} \), \( \text{jinc}(x) \equiv 2J_1(x)/x \), and \( J_1(x) \) is the Bessel function of the first kind. The factor \( \pi d^2/4 \) refers to the area of the aperture. Propagation through the first phase mask results in the field being multiplied by

\[
\left[ 1 - 2\text{circ} \left( \frac{\rho}{s} \right) \right],
\]
i.e. a function which is equal to $-1$ inside a spot of radius $s$ and $+1$ elsewhere. If the spot is small compared to the width of the Airy disk then the amplitude of light within the phase diffracting spot can be approximated as uniform, and hence we can write the light distribution after the phase mask, $A_2(u, v)$, as

$$A_2(u, v) \approx \frac{1}{i \lambda f_1} \frac{\pi d^2}{4} \left[ \text{jinc} \left( \frac{\pi d \rho}{\lambda f_1} \right) - 2 \text{circ} \left( \frac{\rho}{s} \right) \right]. \quad (6)$$

The light distribution in the plane before the second phase mask, $A_3(x, y)$, can be calculated from a Fourier transform expression analogous to that in equation (3). The first term in equation (6) will transform to produce a scaled version of the original aperture function and the second term will transform to give a jinc function, so the total resultant field is

$$A_3(x, y) = \left( \frac{f_1}{f_2} \right)^2 A_0 \left( \frac{f_1 x}{f_2}, \frac{f_1 y}{f_2} \right) - \frac{2}{\lambda^2 f_2^2} \frac{\pi d^2}{4} \pi s^2 \text{jinc} \left( \frac{2\pi sr}{\lambda f_2} \right). \quad (7)$$

If the peak amplitude of the first term in this expression is twice that of the second, then when the light passes through the second phase mask the resulting function will be quasi-continuous (see figure 2(e)). For this to occur we must choose $s$ such that

$$\frac{1}{2} \left( \frac{f_1}{f_2} \right)^2 = \frac{2}{\lambda^2 f_2^2} \frac{\pi d^2}{4} \pi s^2, \quad (8)$$

which, rearranging, gives

$$s = \frac{1}{\pi} \frac{\lambda f_1}{d} \approx 0.31 \frac{\lambda f_1}{d}. \quad (9)$$

It can be shown that in the 1d situation (for example when using a slit instead of a circular aperture) the corresponding value of $s$ is given by

$$s = \frac{\lambda f_1}{8d}. \quad (10)$$

4. Results of numerical simulation

The focal spot generator was simulated numerically and the spot radius, $s$, for which the sidelobe level was minimized was found to be $0.34 \frac{\lambda f}{d}$ for a 2d circular aperture of diameter $d$. The small discrepancy between this value and the previously derived value of $0.31 \frac{\lambda f}{d}$ can be attributed to the fact that the intensity is not completely uniform across the first phase spot. The focal-plane intensity distribution when the optimal spot radius is used is shown in figure 3, and it can be seen that the subsidiary maxima are reduced to a level less than $10^{-4}$ of that of the peak.

5. Limitations

There are a number of limitations to the device proposed here:

1. The above simulation assumes that the extent of the second two lenses and phase mask is infinite. Clearly they must be finite in size, but as long as they
are chosen such that the field due to the jinc function is small at their edges, then this will have a small effect on the final image.

2. The system is critically dependent on the size and phase shifts of the phase masks.

3. For a coronagraph, the alignment accuracy is very critical: adaptive optics or a space telescope is also required. (These comments are also true for other types of coronagraph.)

6. Conclusions

In summary, a series of two phase filters has been proposed which can shape beam to produce an appropriate field distribution so that when focused the resulting image closely approximates a uniform spot with very small subsidiary maxima. Such a device could have a number of applications, for example coronagraphy.

References
Focal spot generator without sidelobes