Realising superoscillations: A review of mathematical tools and their application

To cite this article: K S Rogers and E T F Rogers 2020 J. Phys. Photonics 2 042004

View the article online for updates and enhancements.
TOPICAL REVIEW

Realising superoscillations: A review of mathematical tools and their application

K S Rogers and E T F Rogers

1 School of Mathematics and Statistics, Open University, Milton Keynes, MK7 6AA, United Kingdom
2 Optoelectronics Research Centre and Institute for Life Sciences, University of Southampton, Southampton, SO17 1BJ, United Kingdom

E-mail: k.s.rogers@open.ac.uk

Keywords: superoscillations, imaging, focussing, optimisation, nanophotonics, numerical methods

Abstract

Superoscillations are making a growing impact on an ever-increasing number of real-world applications, as early theoretical analysis has evolved into wide experimental realisation. This is particularly true in optics: the first application area to have extensively embraced superoscillations, with much recent growth. This review provides a tool for anyone planning to expand the boundaries in an application where superoscillations have already been used, or to apply superoscillations to a new application. By reviewing the mathematical methods for constructing superoscillations, including their considerations and capabilities, we lay out the options for anyone wanting to construct a device that uses superoscillations. Superoscillations have inherent trade-offs: as the size of spot reduces, its relative intensity decreases as high-energy sidebands appear. Different methods provide solutions for optimising different aspects of these trade-offs, to suit different purposes. Despite numerous technological ways of realising superoscillations, the mathematical methods can be categorised into three approaches: direct design of superoscillatory functions, design of pupil filters and design of superoscillatory lenses. This categorisation, based on mathematical methods, is used to highlight the transferability of methods between applications. It also highlights areas for future theoretical development to enable the scientific and technological boundaries to be pushed even further in real-world applications.

1. Introduction

Superoscillatory functions, at their most basic, are band-limited functions with the counter-intuitive property that they locally oscillate faster than their highest Fourier frequency. Underneath this simple definition, however, lies a wealth of interesting mathematical properties and practical applications.

Perhaps most surprising is that they can oscillate arbitrarily faster than their highest Fourier frequency over arbitrarily long intervals. That is, a physically band-limited system (such as an optical field, or an electron wavefunction) can locally contain frequencies far outside the highest frequency in its Fourier transform, and what’s more, these oscillations can extend over a wide area. As an extreme example, Michael Berry famously derived a recipe for reproducing Beethoven’s ninth symphony exactly in a signal band-limited to 1 Hz [1].

Berry first used the term ‘superoscillations’ in 1994 [1] referring to work by Aharonov in the context of quantum mechanics [2, 3]. However, functions with this property had already been studied for some time in the mathematics of prolate spheroidal wavefunctions (PSWFs) [4, 5], band-limited functions [6, 7], and oversampling in signal processing (as noted by Berry [1]): reconstructing an oversampled signal (sampled faster than the Nyquist rate) is highly unstable outside the sampled interval [8].

As far back as 1952, Toraldo di Francia [9] was describing optical effects that would now be considered superoscillatory. As such, di Francia was also the first to apply these ideas to optics, giving one of the first applications of superoscillations: focussing light into a sub-diffraction-limited spot. Indeed, applications of superoscillations have interested researchers since the birth of the field: in some of the earliest work on the mathematics of superoscillations, Berry suggests the application of superoscillatory functions of two
Figure 1. Parameters of a superoscillatory spot. Superoscillatory spot (red line - parameters in red text) compared with an Airy disc (black dotted line). The width of the superoscillatory spot, $w$, is usually measured as the full width at half maximum (FWHM) of the central spot, the field of view, FoV, is the distance between the sidebands (both the half-distance/radius and full-distance/diameter are used in the literature). In the pupil-filter literature, the measure [full/half] width at first null is used, shown here for both the superoscillatory focus $W_{so}$ and the Airy disc $W_A$, with their ratio giving the transverse gain $G_T$. A similar axial gain is also used (not shown). Intensity measures include the Strehl ratio $S$, the relative intensity of the sidebands $I_{rel}$, the relative intensity within the FoV $I_{in}$, and the yield $Y$: the ratio of energy in the superoscillatory region to the total energy of the function.

variables to imaging [1], where he notes the possibility of using superoscillatory functions to describe detail much finer than the wavelength of the light used for a new type of super-resolution microscopy. Since then, superoscillations have found many potential applications, both within and outside the field of optics, a selection of which are reviewed in section 6.

Early on in the development of superoscillations, it was also realised that they occur naturally: Berry predicted the existence of superoscillations in random functions in 1994 [1] and the first experimental demonstration to report optical superoscillations was in diffraction from a (quasi) random nanohole array [10]. Concurrent with Berry’s first realisation in 2006 that optical superoscillations can be created in the far field by propagating waves [11], Huang et al [10] showed that a quasicrystalline nanohole array (Penrose-like pattern) can focus light into subwavelength ‘hotspots’ beyond the near field and without the contribution of evanescent waves. Shortly afterwards, the same authors showed that these spots can be projected to the far field [12] demonstrating that the subwavelength features are indeed constructed from propagating waves. Not only has it been shown that random fields contain superoscillations, but also that they are common: on average 1/3 of the area of a speckle pattern (in the scalar approximation) is superoscillatory [13, 14], although it has recently been been shown that a vectorial analysis gives a reduced probability [15]. Konrad and Roux have looked at superoscillations from the perspective of scale physics, and find that superoscillations are natural features of many physical equations, and in some ways should not be unexpected [16].

Alongside these random superoscillations, we also see superoscillations ‘accidentally’ formed in structured fields: a higher-order radially-polarised Laguerre-Gaussian beam is inherently superoscillatory [17], and even using a standard Fresnel zone plate to focus a radially-polarised beam may produce a superoscillatory focus [18].

A key factor in the applicability of superoscillations is the ability to construct them in a controlled manner, and also that they come at a cost: any superoscillatory function must necessarily have a portion of its energy outside the superoscillatory region. In terms of optics, this means that a superoscillatory focus cannot contain all the energy of the beam: some must ‘spill out’ of the focal spot into high-intensity sidebands. Common measures used to quantify superoscillations in optics are the size of the central spot, the field of view (FoV) (the distance between the central spot and the first significant sideband), and some measure of energy efficiency, with the precise measure depending on the application. A range of the most common measures is summarised in figure 1.
Table 1. Quantifying the trade-offs in superoscillations. Parameters from figure 1 calculated for each of the spots in figure 2. The expected trade-offs are clearly seen as yield and Strehl ratio decrease with decreasing spot size and increasing FoV. Here FoV is measured as the distance between the central spot and the first significant sideband.

<table>
<thead>
<tr>
<th>FWHM (λ)</th>
<th>FoV (λ)</th>
<th>Transverse Gain G_T</th>
<th>Relative intensity I_rel</th>
<th>I_in</th>
<th>Strehl Ratio S</th>
<th>Yield Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.43</td>
<td>0.50</td>
<td>0.47</td>
<td>1.31</td>
<td>0.13</td>
<td>0.67</td>
<td>0.38</td>
</tr>
<tr>
<td>0.43</td>
<td>1.00</td>
<td>0.46</td>
<td>1.34</td>
<td>0.088</td>
<td>0.23</td>
<td>0.52</td>
</tr>
<tr>
<td>0.43</td>
<td>1.60</td>
<td>0.46</td>
<td>1.33</td>
<td>0.019</td>
<td>0.18</td>
<td>0.52</td>
</tr>
<tr>
<td>0.39</td>
<td>0.50</td>
<td>0.41</td>
<td>1.49</td>
<td>0.28</td>
<td>0.41</td>
<td>0.19</td>
</tr>
<tr>
<td>0.39</td>
<td>1.00</td>
<td>0.40</td>
<td>1.51</td>
<td>0.078</td>
<td>0.25</td>
<td>0.37</td>
</tr>
<tr>
<td>0.39</td>
<td>1.60</td>
<td>0.41</td>
<td>1.48</td>
<td>0.015</td>
<td>0.22</td>
<td>0.29</td>
</tr>
<tr>
<td>0.35</td>
<td>0.50</td>
<td>0.36</td>
<td>1.71</td>
<td>0.61</td>
<td>0.024</td>
<td>0.22</td>
</tr>
<tr>
<td>0.35</td>
<td>1.00</td>
<td>0.37</td>
<td>1.66</td>
<td>0.62</td>
<td>0.21</td>
<td>0.11</td>
</tr>
<tr>
<td>0.35</td>
<td>1.60</td>
<td>0.37</td>
<td>1.65</td>
<td>0.79</td>
<td>0.19</td>
<td>0.064</td>
</tr>
<tr>
<td>0.31</td>
<td>0.50</td>
<td>0.31</td>
<td>1.97</td>
<td>1.3</td>
<td>0.19</td>
<td>0.11</td>
</tr>
<tr>
<td>0.31</td>
<td>1.00</td>
<td>0.33</td>
<td>1.86</td>
<td>5.9</td>
<td>0.19</td>
<td>0.0069</td>
</tr>
<tr>
<td>0.31</td>
<td>1.60</td>
<td>0.33</td>
<td>1.86</td>
<td>21</td>
<td>0.18</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

The spot size may be measured as the full width at half maximum (FWHM) (w in figure 1), or the full (or half) width at first null (W_{SO}). These may be absolute or compared to the equivalent measure for the Airy disc. In the pupil-filter literature the common relative measure used is the transverse gain, G_T = \frac{W}{W_A}, the ratio of the Airy disc width to the spot width, with gains greater than one indicating a reduced spot size. For 3D analysis the similar axial gain, G_A, is used, which compares the widths at first null in the direction of propagation.

Energy efficiency measures include the proportion of energy in the central spot Y, the Strehl ratio S (ratio of intensity of the central spot to the intensity of an Airy disc with the same total energy) and the ratio of peak sideband intensity to spot intensity I_{rel}. Within the FoV, the ratio of peak intensity beyond the first null to spot intensity, I_{in}, is also used.

As might be expected, there is a trade-off in these measures: it is not possible to have a very small spot in both transverse and axial directions, containing most of the energy, and with very low sidebands. The trade-offs have been investigated by many authors for specific cases, and some general patterns have been seen throughout the literature. The energy in the sidebands is found to increase exponentially with the size of the FoV, but only polynomially with spot size (or equivalently with speed of oscillation) [19]. From the point of view of applications, this polynomial relationship poses much smaller challenges than an exponential one: if the focussed energy decreased exponentially as the spot shrank, then vastly higher laser powers would be needed to realise small spots. The degree of the polynomial relationship is of course important, with a degree of 4 having been reported [20]. Still, a polynomial relationship means that, although we need slightly more powerful lasers, the power trade-offs are not prohibitive. A visual representation of these trade-offs in 2D is given later in figure 4 (section 2.4), with associated parameter values provided in table 1.

It has also been shown that the axial and transverse spot size cannot simultaneously be superoscillatory. A spot with both very small axial and transverse dimensions may therefore not be possible. It should, however, be noted that for some applications this may even be an advantage: some applications require a long spot in the axial direction (see also section 6). A fuller discussion of these trade-offs is given in sections 2 and 3.

To enable quantitative comparisons and statements about superoscillations, a precise quantitative definition is required, yet a few different definitions are currently in common usage. A straight-forward and much-used measure in optical focussing (of light of wavelength \( \lambda \)), is a comparison of the superoscillatory spot size to the diffraction-limited spot size. In 2D the diffraction-limited spot size is the width of the central spot of the Airy disc. Measuring spot size by the FWHM thus gives a superoscillation criterion of \( w < \lambda / 2 \), whereas defining spot size by half width at first null gives a superoscillation criterion of \( W_{SO} < 0.61\lambda \). One issue with this definition is that a simple Bessel beam meets these criteria, and so it may be more appropriate to compare to the equivalent Bessel beam, giving a superoscillation criterion of \( w < 0.38\lambda \). Also, while an obvious choice in optics, this definition is less applicable to the theoretical study of more general superoscillatory functions.

Within this theoretical context an alternative definition of superoscillation is to require the local frequency, defined as the magnitude of the gradient of the phase of the wave, to be greater than the band limit of the function. Given a wave \( \psi(r) = \rho(r) \exp(i\phi(r)) \) with wave number \( k_0 \), the oscillation is superoscillatory where the local wave number, \( k(r) = |\nabla \phi(r)| \), is greater than \( k_0 \) (see for example reference [14]). This definition is readily applicable in certain theoretical contexts, yet it is more natural to consider frequency over an interval than at a point. A further definition therefore considers the number of
zero-crossings of a function in a superoscillating interval. For example, given $2N$ zeros in an interval $[-a,a]$ the frequency can be defined as $\omega = \frac{\pi N}{a}$ and compared to the band limit (see for example reference [20]).

Whichever definition of superoscillation one uses, the aforementioned trade-offs remain. With so many different trade-offs to manage, a plethora of different methods for analysing and optimising superoscillations have been developed. This has been approached from numerous different perspectives, ranging from pure mathematical interest to realising application-specific solutions. This review gives an overview of the available mathematical techniques for constructing superoscillations, to lay out the options to anyone wanting to construct a superoscillation device. We focus on results from the optics literature, which are fairly comprehensive as it was the first application area to have extensively embraced superoscillations, with much recent growth.

The mathematical methods for constructing superoscillations fall neatly into the three categories depicted in figure 2: direct construction of a superoscillatory function, design of a pupil filter (PF) and design of a superoscillatory lens (SOL). The first, reviewed in section 2, gives the most generally applicable approach: a pure mathematical construction of a superoscillatory function\(^3\). This is shown leftmost in figure 2 where the superoscillatory function is realised by back-propagating the field numerically, and then using a metalens with continuous amplitude and phase modulation to impose the back-propagated profile on an incident beam [21, 22]. Forward propagation of the light then reconstructs the superoscillatory field. Superoscillatory functions can also be realised by two other classes of device: a pupil filter, which modulates an incident beam in amplitude and/or phase before it is focussed by a conventional lens (centre panel of figure 2), or a superoscillatory lens which again modulates the amplitude and/or phase of the beam, but in this case does both the modulation and focussing in a single, usually planar, structure (rightmost panel).

Other mathematical methods have been developed specifically for the design of PFs or SOLs, and these are reviewed in sections 3 and 4, respectively. These approaches are more restrictive than freely constructing a superoscillatory function, as design decisions are made \textit{a priori} and must be built into the method. However, they are often a practical solution to realising superoscillations, and manufacturing methods are becoming increasingly sophisticated.

This review draws out and categorises the mathematical methods employed in a multitude of different scenarios, to facilitate the transferability of these methods between different applications. Superoscillations have already been realised in a number of different applications, and section 6 gives a flavour of the extent of these. The final section (section 7) includes concluding remarks and comments on the future outlook of the

\(^3\) It is assumed throughout this paper that conditions such as a function being square-integrable and sufficiently smooth are met as necessary for each method.
field. While we focus on mathematical techniques for constructing and optimising superoscillations, to facilitate their application, fuller descriptions of the history, current applications and future of superoscillations can be found in other reviews, for example by Gbur [23], Chen et al [24] and Rogers and Zheludev [25], and the recent roadmap on the field by Berry et al [26].

2. Constructing superoscillatory functions

Many methods for constructing superoscillatory functions have been developed, each motivated by a different purpose. Some early constructions of superoscillatory functions lend themselves well to the mathematical analysis of their general properties, such as the class of integral representations given by Berry [1], which was inspired by a conversation with Aharonov relating to the weak measurement of a quantum variable [3, 27]. Although the focus of this early paper by Berry is primarily theoretical, he already suggests a new type of super-resolution microscopy as a possible application.

Another construction of superoscillatory functions is proposed and analysed by Berry and Popescu [11]. They consider the band-limited function

$$f(x) = (\cos x + ia \sin x)^N, \quad a > 1, N \gg 1,$$

and note that \(f(x) = \exp(iaNx)\) near \(x = 0\). It therefore oscillates with arbitrarily large frequency if \(a\) is made arbitrarily large.

While this function is primarily used to analyse the persistence of initial superoscillations in a quantum wavefunction as it evolves in free space, an important analogy with light passing through a diffraction grating is drawn. The authors theoretically demonstrate super-resolution without evanescent waves, by drawing a parallel with a diffraction grating which produces a diffraction pattern given by a superoscillatory function, where the superoscillations, made up of purely propagating waves, persist beyond the evanescent region. Equation (1) is also later used by other authors [28–30].

A comprehensive mathematical study of superoscillatory functions expressed in the form of equation (1) is given in [31]. For a broader review of the mathematics underpinning super-resolution, of which superoscillations was one emerging approach at the time, see reference [32]. The focus there is on the mathematical properties of band-limited functions and entire functions within the context of super-resolution of linear optical instruments. This complements the mathematics in this current review, which gives an overview of methods for constructing superoscillatory functions for realisation in applications expanding beyond optical super-resolution.

2.1. Superoscillatory functions by interpolation

Ferreira and Kempf approach the analysis of superoscillatory functions from a different perspective, building on results from signal processing [19]. They study superoscillatory functions which interpolate specified points, and so there is some control of the shape. The interpolation function can contain arbitrarily fast oscillations by specifying arbitrarily close interpolation points that oscillate between, for example, −1 and 1. They analyse the energy cost of superoscillations, and derive the much-quoted and reinforced (for example [33, 34]) relation that the total energy of a superoscillatory function grows exponentially with the number of superoscillations, and polynomially with the speed (the reciprocal of the period) of the superoscillation. Note that polynomial growth with the speed of superoscillation is much better than one might expect, and is very promising in terms of overcoming the necessary challenges to exploiting superoscillations in applications.

As well as exploring the energy cost of superoscillations, Ferreira and Kempf give the unique minimum-energy function that interpolates a set of \(N\) arbitrary real numbers at a set \(t = \{t_k\}_{k=1}^N\) of distinct real numbers. This minimum-energy function, which is band-limited to \(\mu/2\) Hz, is given by

$$f(t) = \mu \sum_{k=1}^N a_k \text{sinc}[\mu(t - t_k)],$$

where the coefficients, \(a_k\), are determined from the interpolation points. This minimum-energy function applies a result by Levi [8]. He considers the interpolation of a finite set of points, spaced more closely than the Nyquist rate, to a band-limited function, finding the minimum-energy solution by variational techniques.

A similar analysis of the energy cost of superoscillatory functions is also done within the context of quantum mechanics [35], and a proof is given that a function always exists which interpolates any finite number of points for any fixed band limit [36].

More control of the shape of the superoscillatory function is introduced by Lee and Ferreira [37] by extending the method of Ferreira and Kempf [19] to allowing values of the derivatives of a target.
superoscillatory function to be specified at any finite number of points, in addition to, or instead of, the amplitude. Variational techniques are again applied to find the unique minimum-energy solution that meets such constraints, and this is shown to always be possible for any given bandwidth.

Further methods from communication theory are used by Ferreira et al [38] to develop a technique for constructing superoscillatory functions based on oversampled signal reconstruction. Here, it is shown that the classical sampling formula can also be used to interpolate a finite set of arbitrarily close, evenly-spaced constraints, \( f(kT) = a_k, \ k \in J, \ T < 1 \)

\[
f(t) = \sum_{k \in I} f(k) \text{sinc}(t - k), \tag{3}
\]

where the number of constraints is equal to the number of coefficients to be determined (i.e. \(|I| = |J|\)). As well as giving a method for constructing superoscillatory functions, this also enables a more thorough study to verify earlier claims that the energy cost of superoscillatory functions is strongly linked with the instability of reconstructing oversampled functions.

Since \(|I| = |J|\), i.e. the number of coefficients to be determined is equal to the number of constraints, this method results in a square, linear system of equations, shown to have a unique solution. Lee and Ferreira extend this to allow \(|I| \neq |J|\), i.e. to considering also the over- and under-determined cases [39]. Such problems lend themselves well to a least-squares approach, with the least-squares solution determined by the pseudo-inverse in the usual way. The numerical stability of finding the pseudo-inverse is discussed in detail, and it is noted that the least-squares solution does, of course, not normally satisfy the constraints in the over-determined case.

Under-determined problems have an infinite set of solutions, and the least-squares approach gives the solution with minimum norm, which in this case is equal to the minimum-energy solution. Lee and Ferreira also show [39] that when this more general problem is reduced to the problem in reference [19] (where the minimum-energy solution is found by variational techniques), the two methods do indeed give the same result.

In the case where the number of constraints is equal to the number of coefficients, Lee and Ferreira also compare the use of equations (2) and (3) for constructing superoscillatory functions [40]. The first equation is known to express an ill-conditioned problem [19] (the expression for \(a_k\) involves the inversion of an ill-conditioned matrix for large numbers of superoscillations) whereas equation (3) always leads to a well-conditioned problem. The downside of equation (3) is that its solution is not of minimal energy. However, Lee and Ferreira show that it typically gives solutions of only approximately 10% higher energy while the dynamic range of the coefficients is 4-5 orders of magnitude smaller [40]. Solutions that maximise a form of energy concentration are also found. The conditioning of this method is also analysed further, with a method for finding the optimal condition number [41].

Katzav and Schwartz recognised that the proportion of energy in the superoscillatory region of a function may be of greater concern in applications than the total energy [20]. They therefore maximise the energy yield of a periodic superoscillatory function, defined as

\[
Y = \frac{\int_{-\pi}^{\pi} f^2(t) \, dt}{\int_{-\pi}^{\pi} f^2(t) \, dt}, \tag{4}
\]

where \([-a, a]\) is the superoscillatory region contained within one period of length \(2\pi\), and \(a < \pi\). By specifying \(M\) interpolation points in \([0, a]\), the shape of an even target superoscillatory function with \(M - 1\) oscillations is controlled in a similar way to above, where the superoscillations can again be arbitrarily fast. These constraints are applied to a periodic function in the form of a truncated Fourier cosine series

\[
f(t) = \frac{A_0}{\sqrt{2\pi}} + \sum_{k=1}^{N} \frac{A_k}{\sqrt{\pi}} \cos(kt), \tag{5}
\]

resulting in a system of \(M\) linear equations in the \(N + 1\) unknowns, \(A_k\), which has an infinite set of solutions in the under-determined case (i.e. when \(M < N + 1\)). The problem of maximising the energy yield is expressed as a generalised eigenvalue problem, which is solved numerically to give the function with optimal energy yield (corresponding to the generalised eigenvector with the largest generalised eigenvalue).

Katzav and Schwartz also extend this analysis to an investigation of the other eigenvectors (of lower energy yield) that are constructed within the method. Sorting the full set of eigenvectors by descending eigenvalue (i.e. by descending energy yield), each eigenvector introduces one more oscillation than the previous one, with the first eigenvector having \(M - 1\) oscillations. This is therefore another method for constructing superoscillatory functions of optimal energy yield where the required oscillations are only
partly pre-specified: the interpolation points prescribe $M - 1$ oscillations with the further eigenvectors providing solutions with naturally arising additional oscillations. As fewer constraints are specified, the energy yield is greater for a given number of oscillations than when the full set of oscillations is prescribed. However, there is a trade-off with the quality of the superoscillations as the additional oscillations are, of course, not constrained.

Taking this process to the limit and specifying only one point gives rise to a yield-optimised method of finding superoscillatory functions with any number of oscillations and effectively without any interpolation. As these solutions are yield-optimised they also enable the study of energy bounds and energy cost as mentioned in the introduction, reinforcing and expanding on the previous results of Ferreira and Kempf [19]. This includes demonstrating an order 4 polynomial relationship between the energy yield and the size of superoscillations, where the size of oscillation is reduced by reducing the size of the superoscillatory region, $a$, while keeping the number of oscillations fixed.

The superoscillatory functions resulting from the methods of Ferreira and Kempf [19] and Katzav and Schwartz [20] are compared in figure 3. Here, the minimum-total-energy and maximum-energy-yield solutions are plotted together for the same interpolation points and band limit. It is reassuring to see how closely they match in the superoscillatory region, and the sidebands differ as expected. The non-periodic and periodic nature of the two functions is clearly seen.

Also controlling the shape of the superoscillatory function by interpolation, but now in 2D, Makris and Psaltis [42] combine this with maintaining a superoscillatory spot on propagation. This is important in some applications, for example to achieve good depth of focus in imaging. They construct superoscillatory functions by using linear combinations of diffraction-free beams, which are well known for their invariance on propagation [43]. A number of interpolation points are chosen, spaced sufficiently closely to give a superoscillatory pattern in the $x$-$y$ plane (which need not be radially symmetric). Applying these constraints to a linear combination of the same number of (diffraction-free) Bessel beams of different order results in a square, linear system of equations which determines the superposition coefficients. No optimisation is
done—this is primarily a demonstration that diffraction-free beams can be superimposed to create superoscillatory patterns that persist on propagation, and without the need for any focussing. The analysis is later extended to a full vectorial treatment [44]. This method of superimposing diffraction-free beams is also extended to create any shape of superoscillatory function [45], with further detail given in section 2.3.

2.2. Superoscillatory functions by manipulating zeros

Superoscillatory functions can also be controlled by specifying the zeros of a band-limited function, rather than interpolating a set of points. A finite number of arbitrarily close zeros, specified for a band-limited function, results in a superoscillatory function. Analysis on which some of these procedures relies is given by Requicha [46], and an early method is described by Qiao [47]. By writing a finite-energy band-limited function as its product expansion, the first $N$ zeros can be modified at will without changing the band limit. The sinc function, for example,

$$\frac{\sin \pi x}{\pi x} = \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{n^2}\right),$$

(6)

can be modified to

$$f(x) = \prod_{n=1}^{N} \left(1 - \frac{k_n^2 x^2}{n^2}\right) \prod_{n=N+1}^{\infty} \left(1 - \frac{x^2}{n^2}\right), \quad k > 0,$$

(7)

which is band-limited to 1/2, yet contains local oscillations of (arbitrary) frequency $k/2$.

An alternative way of designing periodic superoscillatory functions by manipulating zeros is to subtract an amount $s$ ($0 < s < 1$) from a cosine function of frequency $f_0$ [48, 49]

$$f(x) = [\cos(2\pi f_0 x) - s]^2.$$  

(8)

The central peak becomes arbitrarily small as $s$ approaches 1, yet the function remains band-limited. Further zeros can be introduced in the superoscillatory region by iteratively repeating another vertical translation with $s = 0.5$, scaling and squaring, giving a number $2^k$, $k = 1, 2, 3, \ldots$ of zeros in the superoscillatory region. While these further zeros, as well as the shape of the final function, are restricted by the properties of the cosine function, the method is simple and ensures a very smooth and regular result. This method has been applied to structured illumination microscopy [49].

Chojnacki and Kempf construct superoscillatory functions as a product of band-limited functions [34], where each function is chosen to contribute at least one zero. The band limit of these superoscillatory functions is the sum of the band limit of each of the contributing functions. Examples are given for periodic functions (the product of translated sine functions) and square-integrable functions (the product of translated sinc functions). This is a numerically stable method, and the dynamic range in the examples given is shown to be within one order of magnitude of minimum-energy solutions. It is worth noting that while the summation of functions representing waves is straight-forward to realise in optics, the product of such functions is more challenging. However, multiplication of signals is easy and common in electronic systems, such as those used in the generation of radio-frequency superoscillations.

A very different method using the product of functions to obtain superoscillatory functions is given in references [33, 50, 51]. Chremmos and Fikioris note how the well-known property of the Fourier transform

$$\mathcal{F}\{x^n f(x)\} = i^n \frac{d^n}{dk^n} f(k),$$

(9)

can be used to convert any polynomial to a band-limited function by multiplication with a band-limited envelope function, $f(x)$. They use this to construct superoscillatory functions of any shape [33], which we return to in section 2.3. Maintaining the focus on manipulating zeros for now, Smith and Gbur extend this idea to constructing superoscillatory functions in 2D by specifying complex polynomials with arbitrarily close zeros in the form of optical vortices. They demonstrate the method with two examples using real, 2D band-limited envelope functions with circular and rectangular shape in frequency space.

The method is developed further by the same authors, including a full design process for a SOL, as well as simulations demonstrating its focussing performance [51]. Zero rings are now specified, rather than the previous zero points, by considering polynomials in the radial coordinate. Experimental results based on a modification of this method are presented by Lin et al who also consider the propagation of the superoscillations [52].

A similar idea is developed by Mansuripur and Jakobsen [53], where a polynomial with closely spaced zeros is specified and multiplied by a band-limited envelope function. As above, this results in a
superoscillatory function due to the property of the Fourier transform in equation (9). Here, however, a method for controlling a general envelope function is also given. The envelope function is based on a band-limited function \( g(t) \), and the band limit and shape of the final envelope function is controlled using \( g^s(t/\nu) \), where \( \nu \) is a large positive integer. The scaling \( t/\nu \) ensures that the band limit of the original function \( g(t) \) is maintained. As well as introducing a band limit, the envelope function moderates the growth of the sidebands. Examples of multiple possible envelope functions are given, including the possibility of adding or removing zeros from the envelope function, and a full example of a superoscillatory function constructed in this way using a truncated (polynomial) Euler expansion of a cosine function is also given.

### 2.3. Superoscillatory functions with continuously specified shape

As mentioned above, it is possible to construct superoscillatory functions with fully prescribed shape, introducing a completely new level of control over interpolating points or specifying zeros. Within a finite FoV, a superoscillatory function can recreate the shape of any square-integrable function, including functions that are not band-limited and approach the Dirac delta function.

The prolate spheroidal wave functions (PSWFs) \([4, 54, 55]\) have the useful property of providing a double basis set for square-integrable band-limited functions. They are band-limited and orthonormal on the real line, and also form a complete orthogonal set on a finite interval. Thus any square-integrable function can be reconstructed in a finite interval by a linear combination of PSWFs. This is used by Huang and Zheludev to design a transmission function giving a superoscillating diffraction pattern that approximates an arbitrarily narrow sinc function within a FoV \([56]\). Its use for imaging with theoretically unlimited resolution is also described.

In 2D, the radially symmetric circular prolate spheroidal wave functions (CPSWFs) \([5, 57]\) provide an equivalent basis set, which is used by Rogers \emph{et al} to extend the above result to 2D \([58]\). While these results are interesting from a theoretical perspective, the authors note that this method is unlikely to be practical for applications: specifying the precise shape of a superoscillatory function comes at a high cost in terms of the intensity in the sidebands, and the precise shape is unlikely to be of practical importance.

Other basis sets which are simpler and easier to implement can also be used to construct superoscillatory functions of any prescribed shape. Returning to Chremmos and Fikioris \([33]\), the (polynomial) Taylor expansion of a required fast-oscillating function can be made superoscillatory by multiplication with a band-limited envelope function. This relies on the property of the Fourier transform in equation (9), and approximates the required function in a region where the envelope function is close to 1. The following specific example is analysed

\[
f(t) = \frac{3\sqrt{3}}{2} \left( \frac{t^3}{s^3} - \frac{t}{s} \right) \sin^4 \left( \frac{t}{4D} \right),
\]

where \( D \geq 1 \) and \( s < 1 \) are parameters used to control the region where the envelope function approximates 1 and the superoscillatory region, respectively. This method is also modified with a focus on manipulating zeros in 2D \([50–52]\), as described in section 2.2.

Also mentioned above (in section 2.1) was the use of superpositions of diffraction-free Bessel beams of different order, to interpolate specified points \([42]\). The use of Bessel beams extends the axial persistence of the superoscillations. This method is developed further by Greenfield \emph{et al} to an analytic method for generating arbitrarily-shaped superoscillations from the superposition of Bessel beams of different order (and potentially shifted from one another) \([45]\). A target superoscillatory function is approximated by its Taylor expansion, and the electric field resulting from the superposition of Bessel beams is also approximated as a polynomial. The coefficients of the polynomials are then matched to determine the required values of the coefficients of the Bessel beams. Higher order Bessel beams are necessary, as the \( n \)th order Bessel function is needed to match the \( n \)th order polynomial term of the Taylor expansion. As well as giving analytic examples, the propagation of a superoscillation feature constructed by this method is experimentally demonstrated to persist for 250 Rayleigh lengths.

This theme of constructing superoscillatory functions from functions with desirable properties continues as Eliezer and Bahabad create superoscillatory patterns from the interference of Airy beams \([28]\). A method they derived earlier \([59]\) is used to write the imaginary part of equation (1) as a sum of sine modes, providing a target superoscillatory function. These superoscillatory functions can be adjusted by varying the parameter \( a \) in equation (1). Using finite Airy functions as a basis, they also approximate a sum of Airy functions as a sum of sine modes, allowing coefficients and frequencies to be matched and the target function to be reconstructed as a sum of Airy beams. These superoscillatory Airy patterns are also demonstrated experimentally.

Following a similar technique, a target superoscillatory function is specified by Zacharias \emph{et al} by expanding the real part of equation (1) as a sum of cosine modes \([30]\). The interference pattern in the axial
direction is written as a zero-order Bessel beam in the radial coordinate, multiplied by a propagation term, and is also approximated as a sum of cosine modes. This allows coefficients and frequencies to be matched, with the resulting coefficients corresponding to zero-order Bessel beams at different radii. Thus the necessary amplitude and phase profile can be found to realise the target function as the axial profile of a focussed beam of light. Experimental results demonstrate axial superoscillatory focussing by this construction method.

2.4. Specifying properties such as FWHM, \( I_{\text{opt}} \) and FoV

The methods reviewed above construct superoscillatory functions with prescribed shape in some form. For applications, it may be more useful to specify certain properties that a superoscillatory function should have, such as a small FWHM, a large FoV and/or small \( I_{\text{opt}} \) (i.e. low intensity sidebands as depicted in figure 1).

Such considerations date back to the work on supergain antennas by Schelkunoff [60] and Dolph [61], who minimised spot size for a given \( I_{\text{in}} \). Di Francia [9] noted how their results could be applied to the design of PFs to achieve super-resolution in optical focussing, and that the diffraction limit was a practical rather than theoretical limit. A large body of work within the design of PFs is relevant to the use of superoscillations in optical focussing. Here we review the aspect of this early work relating directly to methods for constructing superoscillatory functions, leaving PF design to section 3.

Wong and Eleftheriades brought these methods from antenna arrays into the field of optical superoscillations in a series of results [62–66], where they demonstrate how the methods of Schelkunoff [60] and Dolph [61] (see also [67, 68]) can be used to construct a periodic superoscillatory function, and how that can be realised in applications.

The method relies on the observation [60, 62] that a diffraction pattern from \( N \) uniformly-spaced point sources can be written as a complex polynomial in \( w = e^{-iw \Delta k} \), where \( \Delta k \) is the relative phase shift between each point source. Furthermore, this polynomial can be written in terms of its zeros, \( w_{zn} \),

\[
E(x) = \sum_{n=0}^{N-1} a_n w^n = \prod_{n=1}^{N-1} (w - w_{zn}),
\]

where the \( a_n \) are determined from the specified zeros. Schelkunoff showed that within antenna theory, the zeros can be chosen to design supergain antennas [62], with the direct consequence that the zeros can also be chosen to design superoscillatory functions.

The question of how to optimally choose these zeros was tackled by Dolph [61]. By specifying \( I_{\text{in}} \), the required ratio between the intensity of the central peak and uniform amplitude of ripples between the specified zeros, the optimal spot width is given by a linear combination of Chebyshev polynomials. Equivalently, a required spot width can be specified and a linear combination of Chebyshev polynomials then gives the optimal \( I_{\text{in}} \) (see also [68]).

Thus for a given set of uniformly-spaced point sources and required spot size (or \( I_{\text{in}} \)), the zeros \( w_{zn} \) that result in the optimal \( I_{\text{in}} \) (or spot width) can be determined. From this, the coefficients, \( a_n \), of equation (11) can be found and simple back-propagation using angular spectrum theory determines the amplitude and phase required from the point sources to realise the superoscillatory function as a diffraction pattern [62].

This method is adapted and applied by Wong and Eleftheriades in a number of progressive results. Focussing performance in free space and in a waveguide are demonstrated by simulations [62], with experimental results following for a waveguide [63]. They then extend the method to 2D and build a working microscope [65]. As the specified zeros essentially control the size of the FoV, the method lends itself well to this non-scanning microscope, where objects can be imaged within the FoV. The method is extended further to constructing a broadband spot in all three dimensions, with its potential use for 3D super-resolution imaging also outlined [66].

An interesting new perspective on this method is also presented, where the ripples beyond the (now diffraction-limited) central spot are designed to be superoscillatory, allowing smaller sidebands [69]. Their use for imaging within the FoV is simulated and demonstrated experimentally.

A very different approach to constructing superoscillatory functions with optimal properties is offered by optical eigenmodes, introduced and defined by Mazilu et al [70]. Optical eigenmodes provide a natural basis for finding the optimal spot size for a given FoV (referred to as region of interest by the authors) and intensity threshold. They arise from the linearity of Maxwell’s equations combined with the observation that many measures of interest, such as energy and intensity, are quadratic in nature. The theory is fully developed for vector fields, but scalar fields are later applied to optical imaging [71].

Consider a scalar electric field written as the superposition of \( N \) electric fields, \( E = \sum_{k=1}^{N} a_k E_k \). The intensity (or any other quadratic measure) integrated over a FoV may then be written in matrix form, where each element of a matrix, \( \mathbf{M}^{2D} \), is given by
for three different shows these trade-offs in detail, defining the spot width to be applied to imaging of biological samples. Low-intensity sidebands (which are not forbidden inside the FoV) are dominant. These can move from and increasing FoV. The relative intensity, $I_{rel}$, tends to increase, but is less predictable for large spots, where low-intensity sidebands (which are not forbidden inside the FoV) are dominant. These can move from outside to inside the FoV as it increases in size. The semi-analytic method is sufficiently fast to be re-calculated in real-time and has been realised as a superoscillatory intensity pattern using an SLM and applied to imaging of biological samples.

This method is also used to design a metalens [74]. Metalens technology such as this, with (nearly) continuous amplitude and phase modulation, opens up the opportunity of easily realising any directly-designed superoscillatory function, removing any constraints imposed by PF and SOL design. The downside is that these metalenses require complex nanofabrication technologies to create.

The choice of CPSWFs as a basis set enabled independent control of the constructed superoscillatory function within the FoV. Different basis sets can be chosen with different advantages. Inspired by Wong and Eleftheriades [62], the Chebyshev polynomials are used by Xie et al as a natural choice for maintaining a low intensity beyond the central spot within a large FoV [75, 76]. Here, a GA is implemented with the intensity in the focal plane written as a complex combination of Chebyshev polynomials. The objective is to maximise the intensity of the central peak for a pre-specified spot size, $I_0$, and FoV. These restrictions are imposed as specified zeros, with the first zero determining the spot size, the FoV determined by the interval over which the zeros are specified and $I_0$ given by a threshold.

\[
M_{jk}^{10} = \int_{\text{FoV}} E_j^* E_k 2\pi dr,
\]

where $^*$ denotes complex conjugation and radial symmetry has been assumed (though this is not necessary). For a specific superposition of the $E_k$, given by the vector $a$ with elements $a_k$, the intensity is then $a^\dagger M^{10} a$, where $a^\dagger$ is the conjugate transpose of $a$.

The eigenvectors of the matrix, $M^{10}$, specify a superposition of the initial set of electric fields with particularly useful properties, and these eigenvectors are defined to be the optical eigenmodes. For example, the eigenmode corresponding to the largest eigenvalue of $M^{10}$ gives the superposition of the initial set of electric fields, $E_k$, for which the intensity in the FoV is maximal, with the intensity given by that largest eigenvalue.

An equivalent definition applies to optical eigenmodes for any other quadratic operator. Spot size can also be given by a quadratic measure, by using the second moment of the intensity giving a matrix $M^{2\alpha}$, and defining the spot width to be

\[
w = 2 \sqrt{\frac{a^\dagger M^{2\alpha} a}{a^\dagger M^{10} a}}.
\]

Using the normalised eigenmodes from $M^{10}$ as the input fields, $E_k$, to find the eigenmodes of $M^{2\alpha}$, results in the minimum spot of maximal intensity being given by the eigenmode of $M^{2\alpha}$ with smallest eigenvalue. Only eigenmodes of $M^{10}$ that meet a specified intensity threshold are used.

As almost all optical systems are linear, this analysis can be applied to input fields in the input plane with measures taken on the resulting fields in the output plane. No knowledge of any complex optical system between the input field and the output is needed. Optimisation of the output parameters then gives the optimum superposition of input fields. Both simulations and experimental results are given, demonstrating the use of this method to create superoscillatory spots using a spatial light modulator (SLM) to encode the required superposition of eigenmodes. The method is further applied to experimentally focussing light into superoscillatory spots in the far field [72], and a complete confocal scanning imaging system [71].

As we have seen, there is a trade-off between several measures of interest. Ideally, we would simultaneously optimise the FWHM, $I_{rel}$, and FoV, to name but a few. However, improving any one of these causes a deterioration in another. Rogers et al implement a genetic algorithm (GA) to perform a multi-objective optimisation which simultaneously minimises the FWHM and $I_{rel}$ for a number of different FoV, obtaining superoscillatory functions with the best compromises (a Pareto front) [58]. The problem space is all possible superoscillatory functions, which is approximated by linear combinations of CPSWFs, and so the population for the GA is a set of functions. More detail is given about GAs in section 4.3, where they are used in numerous cases to optimise SOLs.

Based on the results from this GA, Rogers et al also propose a much simpler and faster semi-analytic method. A linear combination of only two CPSWFs, selected for their natural spot size, is shown to give similar results. Twelve best compromises obtained using this method are shown in figure 4 for three different FoV, illustrating clearly the inevitable trade-offs of superoscillations. As spot size decreases (vertically), and as FoV increases (horizontally), the intensity in the sidebands increases. Table 1 shows these trade-offs in detail, with the same patterns as in the figure. In all cases, yield and Strehl ratio decrease with decreasing spot size and increasing FoV. The relative intensity, $I_{rel}$, tends to increase, but is less predictable for large spots, where low-intensity sidebands (which are not forbidden inside the FoV) are dominant. These can move from outside to inside the FoV as it increases in size. The semi-analytic method is sufficiently fast to be re-calculated in real-time and has been realised as a superoscillatory intensity pattern using an SLM and applied to imaging of biological samples [73].

This method is also used to design a metalens [74]. Metalens technology such as this, with (nearly) continuous amplitude and phase modulation, opens up the opportunity of easily realising any directly-designed superoscillatory function, removing any constraints imposed by PF and SOL design. The downside is that these metalenses require complex nanofabrication technologies to create.

The choice of CPSWFs as a basis set enabled independent control of the constructed superoscillatory function within the FoV. Different basis sets can be chosen with different advantages. Inspired by Wong and Eleftheriades [62], the Chebyshev polynomials are used by Xie et al as a natural choice for maintaining a low intensity beyond the central spot within a large FoV [75, 76]. Here, a GA is implemented with the intensity in the focal plane written as a complex combination of Chebyshev polynomials. The objective is to maximise the intensity of the central peak for a pre-specified spot size, $I_0$, and FoV. These restrictions are imposed as specified zeros, with the first zero determining the spot size, the FoV determined by the interval over which the zeros are specified and $I_0$ given by a threshold.
Figure 4. Visualising the trade-offs in optimised superoscillations. Twelve best compromises are shown, between spot size and $I_{\text{rel}}$, obtained using the semi-analytic optimisation method of Rogers et al [58]. This illustrates clearly the inevitable trade-offs of superoscillations: as spot size decreases (vertical axis), and as FoV increases (horizontal axis), the intensity in the sidebands increases. Here, the FoV is measured as the distance between the central spot and the first significant sideband. These trade-offs are quantified in table 1.

A binary phase PF with $N$ zones is used to realise this superoscillatory function. The intensity profile in the focal plane of the PF is given by the Hankel transform of a transmission function taking values 1 and -1, to impose phase shifts of 0 and $\pi$ (assuming the Fraunhofer approximation). This expression depends on the position of the radii of the zone boundaries, and also forms part of the optimisation by minimising the mean squared error between the intensity profile in the focal plane written as a combination of Chebyshev polynomials and as a sum of the contributions from each zone of the binary PF. The method thus combines the construction of a superoscillatory function with the design of a PF. To increase the depth of focus, a pre-specified central part of the PF is also obscured, and a working non-scanning microscope designed using this method is built and demonstrated [76].

Before we move on to considering the explicit design of PFs, we provide a summary of the above methods for the direct design of superoscillatory functions in table 2. Each method is briefly characterised by its main features, such as whether it constructs 1D, 2D or 3D, single- or multi-lobe, or any shape of superoscillatory functions, whether it uses scalar or vector theory and whether any application to focussing or imaging is included. A brief comment on the purpose of each method is also included, with the full table providing an at-a-glance overview of the many available methods.

3. Direct design of pupil filters

Starting with the aim of designing a PF to achieve super-resolution focussing, as opposed to starting with the aim of designing a superoscillatory function, approaches the same problem from the opposite side. Once a superoscillatory function is constructed, for example by one of the methods reviewed above (in section 2), the problem of realising it as an electric field remains (which can be done in various ways). Starting instead
Table 2. A summary of mathematical methods for constructing superoscillatory functions. Each method is characterised as follows: 1D, 2D (normally radially symmetric) or 3D; consideration of any axial elongation (E) or compression (C); use of scalar (S) or vector (V) analysis; resulting in single- (S) or multi-lobe (M) functions, or functions of any shape (A). Any restrictions on the shape of function are also summarised, with a brief comment on the main purpose of each method. The final column indicates any application to focusing or imaging by simulation (S) and/or experiment (E).

<table>
<thead>
<tr>
<th>reference</th>
<th>1D/2D/3D</th>
<th>Axial elongation/compression</th>
<th>Scalar/vector</th>
<th>Single/multi lobe/Any shape</th>
<th>Restrictions on shape</th>
<th>Purpose</th>
<th>Applied to focusing/imaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>[11]</td>
<td>1D</td>
<td>-</td>
<td>S</td>
<td>M</td>
<td>Particular construction</td>
<td>Demonstrate persistence of superoscillations beyond the evanescent region</td>
<td>-</td>
</tr>
<tr>
<td>[19, 37]</td>
<td>1D</td>
<td>-</td>
<td>S</td>
<td>M</td>
<td>Amplitude and/or derivative constraints</td>
<td>Analyse energy cost, minimum-total-energy solution</td>
<td>-</td>
</tr>
<tr>
<td>[38–40]</td>
<td>1D</td>
<td>-</td>
<td>S</td>
<td>M</td>
<td>Amplitude constraints, uses classical sampling formula</td>
<td>Optimise energy concentration</td>
<td>-</td>
</tr>
<tr>
<td>[42, 44]</td>
<td>2D</td>
<td>E</td>
<td>S[42]/V[44]</td>
<td>M</td>
<td>Amplitude constraints</td>
<td>Diffraction free</td>
<td>-</td>
</tr>
<tr>
<td>[47]</td>
<td>1D</td>
<td>-</td>
<td>S</td>
<td>M</td>
<td>Specified zeros, particular band-limited function</td>
<td>Method using control of zeros</td>
<td>-</td>
</tr>
<tr>
<td>[48, 49]</td>
<td>1D</td>
<td>-</td>
<td>S</td>
<td>M</td>
<td>Use of cosine function</td>
<td>Smooth and regular shape</td>
<td>S/E</td>
</tr>
<tr>
<td>[34]</td>
<td>1D</td>
<td>-</td>
<td>S</td>
<td>M</td>
<td>Specified zeros, particular band-limited functions</td>
<td>Computationally efficient</td>
<td>-</td>
</tr>
<tr>
<td>[50, 51]</td>
<td>2D</td>
<td>-</td>
<td>S</td>
<td>M</td>
<td>Specified zeros</td>
<td>Superscillations by positioning optical vortices [50] or zeros [51]</td>
<td>S[51]/E[52]</td>
</tr>
<tr>
<td>[53]</td>
<td>1D</td>
<td>-</td>
<td>S</td>
<td>M</td>
<td>Specified zeros, particular band-limited functions</td>
<td>Computationally efficient, control of envelope function</td>
<td>-</td>
</tr>
</tbody>
</table>

*Not constrained to radial symmetry [42].

**Number of lobes must be \( n^2 - 1 \).

*a modified method is used in [32].
<table>
<thead>
<tr>
<th>reference</th>
<th>1D/2D/3D</th>
<th>Axial elongation/compression</th>
<th>Scalar/vector</th>
<th>Single/multi lobe /Any shape</th>
<th>Restrictions on shape</th>
<th>Purpose</th>
<th>Applied to focusing/imaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>[56, 58]</td>
<td>1D [56]/2D [58]</td>
<td>-</td>
<td>S</td>
<td>A</td>
<td>-</td>
<td>Reconstruct any shape</td>
<td>-</td>
</tr>
<tr>
<td>[33]</td>
<td>1D</td>
<td>-</td>
<td>S</td>
<td>A</td>
<td>-</td>
<td>Reconstruct any shape</td>
<td>-</td>
</tr>
<tr>
<td>[45]</td>
<td>2D</td>
<td>E</td>
<td>S</td>
<td>A</td>
<td>-</td>
<td>Reconstruct any shape, diffraction free</td>
<td>E</td>
</tr>
<tr>
<td>[62, 65]</td>
<td>1D [62]/2D [65]</td>
<td>-</td>
<td>S</td>
<td>S</td>
<td>Single peak with uniform ripples in FoV</td>
<td>Optimise spot size for given $I_m$ and FoV</td>
<td>S/E [63, 65]</td>
</tr>
<tr>
<td>[66]</td>
<td>3D</td>
<td>C</td>
<td>S</td>
<td>S</td>
<td>Single peak with uniform ripples in FoV</td>
<td>Optimise 3D spot size for given $I_m$ and FoV $^{10}$</td>
<td>S</td>
</tr>
<tr>
<td>[69]</td>
<td>2D</td>
<td>-</td>
<td>S</td>
<td>M</td>
<td>Uniform ripples</td>
<td>Use of superoscillations beyond diffraction-limited spot to reduce sidebands</td>
<td>S/E</td>
</tr>
<tr>
<td>[70–72]</td>
<td>2D</td>
<td>-</td>
<td>S/V $^{11}$</td>
<td>S</td>
<td>Single spot</td>
<td>Minimise spot size for given FoV and intensity threshold</td>
<td>S/E</td>
</tr>
<tr>
<td>[58] $^{12}$</td>
<td>2D</td>
<td>-</td>
<td>S</td>
<td>S</td>
<td>Single spot</td>
<td>Minimise spot size and $I_{rel}$ given FoV</td>
<td>S/E [73, 74]</td>
</tr>
<tr>
<td>[75, 76]</td>
<td>2D</td>
<td>E</td>
<td>S</td>
<td>S</td>
<td>Specified zeros</td>
<td>Maximise spot intensity given spot size, $I_m$ and FoV</td>
<td>S/E</td>
</tr>
</tbody>
</table>

$^1$ considering only reconstruction by (C)PSWFs.
$^8$ not constrained to radial symmetry.
$^9$ in axial coordinate.
$^{10}$ using broadband waveform.
$^{11}$ imaging is scalar only.
$^{12}$ the GA and semi-analytic approach.
with a PF, the emphasis is on finding a design of the PF that will create an electric field described by a superoscillatory function.

The theoretical design of a superoscillatory function comes with no restrictions; we have already reviewed methods for constructing any superoscillatory function. Although it is useful to consider practical restrictions, this poses no restrictions on the superoscillatory functions that can be designed to meet them. Approaching the problem from the point of view of designing a PF normally does pose restrictions on the realisable electric field profiles.

Most PFs are radially symmetric (if 2D) and discretised into some number of concentric zones of varying width at different radii. Each zone applies certain amplitude and/or phase shifts. Some design decisions have usually been made a priori, for example whether the PF will have an amplitude or phase profile and how many concentric zones it will be made up of (and/or the width of those zones, the radii etc). This must form the starting point of any subsequent mathematical approach, restricting the possible outcomes. The significance of these restrictions, however, depends on the particular PF and can be negligible. As an extreme example, a continuous phase and amplitude PF would pose no such restrictions.

Simpler PFs are often a good practical solution despite any imposed restrictions, and the design of PFs comes with a rich body of research from which much can be learnt about superoscillations. The search for super-resolving PFs without the use of evanescent waves was active long before the term ‘superoscillations’ was introduced. Already in 1952 Di Francia realised that results from supergain antennas could be applied to the design of super-resolving PFs [9], and the more recent development of this by Wong and Eleftheriades to design superoscillatory functions [62] has already been noted. A selection of results from PF design, with an emphasis on relevance to superoscillations, is reviewed here. For a fuller review of PFs, see for example reference [32].

As well as suggesting that Schelkunoff’s [60] method for designing supergain antennas can be used to design PFs, Di Francia also investigates the use of selected Bessel beams for designing super-resolving PFs [9]. He notes the well-known result that a zero-order Bessel beam of the first kind, obtained from a PF transmitting only a thin ring at some diameter, \( D \), results in a smaller focal spot than the diffraction-limited spot. He then suggests adding further such Bessel beams in order to reduce the outer ripples that inadvertently come with Bessel beams at greater radii, essentially increasing the FoV. This is first done by adding two further rings at diameters \( D/3 \) and \( 2D/3 \) with their associated Bessel beams. The coefficients of these Bessel beams is found by imposing an amplitude of 1 at the centre of the spot and zeros at two further positions (the first of these coinciding with the first zero of the original Bessel beam), resulting in a system of three linear equations in the variables. Cases with four and five rings are then shown to push the outer ripples further from the spot, increasing the FoV.

Although Di Francia stops at 5 rings, he notes that the process can be continued indefinitely to achieve an arbitrarily large FoV. He also demonstrates how the spot size can be reduced by choosing the first zero closer to the peak, and notes the cost of a significant increase in the size of the sidebands. This step is important in that the spot size then reduces below that of the equivalent Bessel beam, which has been suggested as an alternative baseline for the claim of superoscillations [77] (as opposed to comparing to the Airy disc, for example).

It is also worth noting that this technique of specifying zeros, the first to set the spot size and the remaining ones to control the FoV, closely resembles techniques for constructing superoscillatory functions in both sections 2 and 4. (Though this selection of zeros is quite different to that used by Wong and Eleftheriades [62], where the relationship between the complex zeros and the spot size and FoV is less direct.) Also, if Bessel beams can be selected at any radius, rather than pre-specifying selected radii, then this essentially provides a complete basis set for the electric field profile in the focal plane, with no additional restrictions imposed by the PF.

Di Francia’s results are based on an entirely theoretical analysis, and a perfect realisation of this PF would require infinitesimally thin rings at the chosen diameters with continuous amplitude and phase modulation. These results are built on in many ways. Di Francia himself proposes a PF with \( n \) finite-width annuli with any amplitude transmittance. For a lens of radius \( a \), the \( n + 1 \) radii of the \( n \) zones are given by \( a\alpha_k, k = 0, \ldots, n \). Here \( \alpha_k = \sqrt{k/n} \) so that the radii are located at \( 0, a\sqrt{1/n}, a\sqrt{2/n}, a\sqrt{3/n}, \ldots, a \) and each zone has equal area of \( \pi a^2/n \). The contribution to the resulting field from the \( k \)-th zone is given by

\[
f_k(v) = (\pi a^2)c_k \left( \frac{2J_1(\alpha_k v)}{\alpha_k v} - \frac{\alpha_k^2 J_1(\alpha_k v)}{\alpha_k v} - \frac{2J_1(\alpha_{k-1} v)}{\alpha_{k-1} v} \right), \tag{14}
\]

where \( J_1 \) is a Bessel function of the first kind, order 1, the \( c_k \) specify the transmittance of the \( k \)-th zone, \( v = 2\pi r \sin \theta / \lambda \), \( r \) is the radial coordinate, \( \lambda \) is the wavelength of the light used and \( \theta \) is the semiangular aperture. The \( c_k \) are determined from a system of \( n \) linear equations resulting from imposing \( n \) independent...
constraints [78]. Cox et al compare this to a number of later proposed theoretical PF designs with continuously varying amplitude [79], and little advantage is found to the harder-to-fabricate continuous amplitude PFs.

Axial focussing behaviour, and its relation to transverse spot size, is considered by Sheppard and Hegedus [80]. As noted in the introduction, it can be desirable either to reduce or to elongate a focal spot in the axial direction, and both cases are considered here. A continuous real-valued (i.e. modulating the amplitude only) PF, \( P(\rho) \), is studied theoretically by writing the intensity in the focal region as

\[
U(v, u) = 2 \int_0^1 P(\rho)J_0(\nu \rho) \exp(i \nu \rho^2/2) \rho d\rho, \quad \text{so that}
\]

\[
U(v, 0) = 2 \int_0^1 P(\rho)J_0(\nu \rho) d\rho,
\]

\[
U(0, u) = 2 \int_0^1 P(\rho) \exp(i \nu \rho^2/2) \rho d\rho,
\]

where \( \rho \) is the radial coordinate in the pupil plane, \( v \) and \( u \) are the radial and axial coordinates in the focal region and \( J_0 \) is the zero-order Bessel function of the first kind. Using a change of variable \( \tau = \rho^2 \), the electric field profile in the focal region is written as a power series and squared to give the intensities

\[
I(v, 0) = I_0^2 - \frac{1}{2} v^2 I_1 + \ldots
\]

\[
I(0, u) = I_0^2 - \frac{u^2}{4} (I_0 I_2 - I_1^2) + \ldots
\]

where \( I_n = \int_0^1 Q(t)^n dt \) is the \( n \)th moment of the pupil function, \( Q(t) \), in terms of \( t \). Only the first two terms (shown above) are retained, and Sheppard and Hegedus note by inspection of these equations that the transverse and axial spot size are inter-related. This inter-relationship is investigated further for a few specific analytic PFs with some experimental verification. In no case do they find that both axial and transverse gain, \( G_A \) and \( G_T \), can simultaneously be greater than one, i.e. the spot is not simultaneously reduced in both transverse and axial directions (see figure 1 for definitions of gain). Sheppard and Hegedus also analyse two specific, discrete amplitude PFs with two and three zones in terms of this inter-relation, and they give two specific continuous amplitude designs for axial focussing and elongation.

Sales and Morris consider phase-only PFs divided into \( N \) concentric zones [81]. The electric field in the focal plane is given by

\[
E(v) = \sum_{k=1}^{N} \exp(i \phi_k) \left( \alpha_k^2 \frac{2J_1(\alpha_k v)}{\alpha_k v} - \alpha_{k-1}^2 \frac{2J_1(\alpha_{k-1} v)}{\alpha_{k-1} v} \right),
\]

where the contribution from the \( k \)-th zone is given by a similar expression to that in equation (14), but with a phase shift \( \phi_k \) instead of an amplitude transmission coefficient. Imposing \( N \) constraints on this system therefore results in a system of \( N \) non-linear equations, rather than the set of linear equations which can be obtained for amplitude modulation (if the radii \( \alpha_k \) are pre-specified). Such a non-linear system is difficult to solve, especially if the \( \alpha_k \) are also treated as variables, and there is no guarantee of a solution. Before considering this inverse problem, Sales and Morris start with an exploratory approach to solving the direct problem for two zones.

Expanding the electric field profile in the focal plane as a power series and retaining the first two terms allows the transverse gain and Strehl ratio to be written in terms of the phase shift, \( \phi_0 \), and radius of the zone boundary, \( \alpha_0 \) (in the case of two zones there is only one value for each). Specifying a required Strehl ratio thus allows the spot size to be plotted as a function of \( \alpha_0 \) for a given phase shift, for example, or the Strehl ratio to be plotted as a function of \( \alpha_0 \) for various values of the phase shift. Imposing an intensity ratio restricts the values of \( \phi_0 \) and \( \alpha_0 \).

The trade-off between spot size, Strehl ratio and intensity ratio can therefore be visualised and optimised by exhaustive exploration of the different values of the parameters. This approach is not practical for more than two zones, so for three zones values from the two-zone PF are used as a starting point.

For more than 3 zones the inverse problem, a set of non-linear equations obtained from a set of constraints, must be solved. This is done by specifying zeros of the electric field profile where, as Di Francia did [9], the first zero determines the spot size and the rest control the FoV. To simplify the problem, a binary phase profile (taking values of 0 and \( \pi \), with \( \phi_0 = \pi \) ) is imposed. A further modification allows control of the Strehl ratio at the expense of controlling the FoV, retaining only some control of the spot size.
Sales and Morris also demonstrate axial focussing in binary phase-only PFs [82]. Considering a two-zone PF designed for use in a confocal scanning microscope, the axial gain is shown to decrease linearly with the phase shift, $\phi_0$. Given a value of $\phi_0$, the optimum zone boundary is shown to occur at $\alpha = \sqrt{1/2}$. The relationships of axial gain, Strehl ratio and intensity ratio, $I_{rel}$, with phase shift are shown, with the trade-off between these values clearly seen, and an improvement in Strehl ratio is gained over amplitude PFs. Sales and Morris also give an example of a binary phase-only five-zone PF, allowing more control over the spot size and FoV than the two-zone PF, while maintaining the improvement in Strehl ratio.

Having considered transverse and axial focussing separately, Sales proceeds to consider simultaneous focussing in three dimensions [83]. Consistent with Sheppard and Hegedus [80], he finds that transverse and axial focussing are inter-dependent, and he also derives a minimum spot size. A focal spot can therefore not be made arbitrarily small in all three dimensions.

A complex PF, allowing simultaneous amplitude and phase modulation, is considered by de Juana et al [84]. Defining $P(\rho) = T(\rho)\exp(i\phi(\rho))$ in equations (15), they extend the analysis for amplitude PFs by Sheppard and Hegedus [80] to this more general case. Making the same approximations, they obtain a complex version of equations (16) with a similar form. These results are used to design a continuous phase-only PF with $\phi(\rho) = a\sin(2\pi b\rho)$. Here, the parameters $a$ and $b$ allow certain properties to be imposed on the desired spot, such as Strehl ratio and intensity ratio. These restrictions result in a system of non-linear equations from which $a$ and $b$ are determined by a numerical method.

To avoid these numerical calculations, a simple, analytic method for designing binary phase-only two-zone PFs is derived by Cagigal et al [85]. The phase is restricted to taking values 0 and $\pi$, leaving the radius of the single zone boundary, $\alpha_0$, as the only free parameter. The intensity profile in the focal region is described by the equations of de Juana et al [84], which reduce to equations (16) for these restricted phase values. This enables the Strehl ratio, transverse and axial gain to be written in terms of $\alpha_0$, and thus the zone boundary to be determined for any required value of these measures. A trade-off between spot size and Strehl ratio is observed, and while axial focussing is considered, it is shown that an axial gain greater than 1 is not possible in this configuration.

Sheppard and Choudhury [86] consider the effect of polarisation on PF design. They describe how properties of plane-polarised, circularly-polarised and radially-polarised light affect transverse and axial spot size, and point out benefits to using radially-polarised light, for example to reduce transverse spot size.

After a move towards phase-only PFs due to their increased transmittance, Sheppard points out that energy loss is not an issue in many applications [87]. He considers a continuous amplitude PF with parabolic transmittance, specifically a Sonine filter, and proves that it maximises a defined signal concentration factor for a given spot size. The axial and 3D focussing of this PF are also described. With superoscillations now emerging strongly in the literature, Sheppard also notes the link between the fields of PF design and superoscillations.

**4. Direct design of SOLs**

PFs engineer the light incident on a lens, with the lens then focussing the light onto a focal plane. However, both these functions can be done in one: a thin mask can be designed to both engineer and focus the light into superoscillatory spots. Here we call such a mask a superoscillatory lens (SOL). These are, as with PFs, normally radially symmetric (if 2D) and discretised into some number of concentric zones of varying width at different radii. Each zone applies a different amplitude and/or phase modulation.

In the same way as taking the design of a PF as a starting point for realising superoscillatory focussing, the starting point may be to design a SOL. Restrictions are similarly imposed by design decisions made a priori, so that the possible superoscillatory pattern in the focal plane, described by a superoscillatory function, is limited by the intended design of the SOL. Such restrictions must form part of any optimisation process. Even if the SOL is a metalsen, then the design and manufacturing techniques of the metalsen may impose some restrictions.

SOLs are a practical and powerful method for constructing superoscillatory spots, and much of the recent progress in realising superoscillations has been in the explicit design of SOLs. The first SOL was demonstrated by Huang et al where a nanohole array based on a Penrose-like pattern was shown to focus light into superoscillatory spots [12]. This section is a review of methods for designing SOLs to realise superoscillatory focal patterns with desirable properties.

**4.1. Direct solution to the inverse problem**

As has been done both to construct superoscillatory functions and PFs, one way of specifying a desired intensity profile is to specify zeros and/or amplitudes. Imposing $N$ restrictions, the first one may be used to specify an amplitude of 1 at the centre, the next to restrict the size of the spot and the remainder to create a
surrounding FoV, for example. This is the approach taken by Huang et al to design an amplitude and phase SOL, where the modulation of a number of finite-width rings is given by a complex number (allowing both amplitude and phase to take any values) [77].

The intensity profile of each ring can be written as an equation of ring radius (at the centre of each ring), ring width and distance to the focal plane. Specifying the value for the width of the rings as \(0.3\lambda\) and the distance to the focal plane as \(20\lambda\) leaves the radii of each ring as the only variable. Imposing \(N\) constraints on a SOL with \(N\) rings then leads to a system of \(N\) non-linear equations. This inverse problem is difficult to solve, and to improve the behaviour of the problem, Huang et al compare the intensity profile from rings with varying width to Bessel functions. The pre-specified width of the rings is chosen such that this agreement is generally good.

A numerical method is required to solve the non-linear system of equations, and a specific example is given with \(N = 5\). The phase shift between rings is found to oscillate between values close to 0 and \(\pi\), and the authors suspect a non-optimal solution caused by numerical problems. A binary phase-only PF (with values 0 and \(\pi\) only) is therefore also considered. This simplifies the equations significantly, and an improved solution is found with higher values of \(N\).

### 4.2. Particle swarm optimisation

Given the difficulties in solving the inverse problem of SOL design as indicated above, it is natural that stochastic global numerical optimisation schemes are widely used. One such method is particle swarm optimisation (PSO). Each potential (randomly generated) SOL design is treated as a particle moving in an \(N\)-dimensional search space, where \(N\) is the number of degrees of freedom in the design. The particles are given an initial random position (equivalent to a design) and velocity, and the fitness of each design is evaluated. In each iteration of the optimisation, the particle position (design) is updated based on its velocity, and its velocity is updated so that it tends to move towards both the best position (design) that particle has found, and the best position (design) found by any particle so far. The algorithm finishes either after a set number of steps, or after the particles have all converged or become static.

Rogers et al use binary particle swarm optimisation to design a binary amplitude SOL with zones of varying width of 0 or 1 amplitude transmittance [88]. The objective of the optimisation is to minimise spot size, and a practical FoV and intensity ratio of the sidebands, \(I_{\text{rel}}\), are included as constraints. Using the angular spectrum method, the electric field profile in the focal plane is found for each potential SOL. Starting with 100 equal-width concentric zones, the final optimised lens has 25 transparent zones of different widths at different radii (neighbouring zones with the same amplitude are combined during the optimisation).

This SOL was created in order to build and demonstrate the first working superoscillatory microscope, and variations on it are later also used. Optical needles are realised by obscuring the central part of the SOL to obtain longer depth of focus and larger FoV [89]. This is a result of the outer zones giving beams that are closer to the diffraction-free Bessel beams. A vectorial treatment using circularly-polarised light in UV is also developed [90].

Achromatic SOLs are designed using particle swarm optimisation by Yuan et al [91]. Both an amplitude SOL with transmittance of 0 or 1, and a phase SOL with 0 or \(\pi\) phase shift, are designed. A target intensity profile is specified for several different wavelengths, and the optimisation algorithm continues until this target is sufficiently well approximated by the SOL configuration that has been reached. Both phase and amplitude SOLs that simultaneously focus two or three different wavelengths into superoscillatory spots are simulated and demonstrated experimentally.

Yang et al use PSO to optimise a binary phase SOL in the THz region for circularly-polarised light [92]. The vectorial angular spectrum method is therefore used to describe the electric field profile in the focal plane (located at \(210\lambda\)). The resulting equations include a transmission function, \(t(r)\), which imposes a phase shift of 0 or \(\pi\) as a continuous function of the radius, \(r\), of the SOL. Three weighted terms specifying FWHM, sideband intensity and peak intensity are incorporated into a single objective to obtain the transmission function that realises these specifications.

PSO is also used to design a supercritical lens array [93], and binary phase SOLs to realise superoscillatory optical hollow spots and needles [94–96] as well as 3D hollow spots [97, 98].

### 4.3. Genetic algorithms

Genetic algorithms are another stochastic global numerical optimisation technique, and have recently been applied in many optimisation problems for SOL design. (As mentioned in section 2.4, these have also been used to design optimal superoscillatory functions.) A genetic algorithm mimics natural evolution, with an initial random population evolving to an optimal one over a number of generations. The initial population can, for example, be a random set of possible SOLs with zones of varying radii, width, amplitude transmittance and/or phase shift. Each SOL in a population is evaluated according to one or more fitness
functions, and given a numerical score. A fitness function can, for example, be defined as the spot size, spot intensity, FoV and/or $I_{\text{rel}}$, as calculated from the intensity profile given by the angular or vectorial spectrum method. The equations describing this intensity profile usually include a transmission function, which specifies the amplitude and/or phase profile of the SOL.

The SOLs in a population are then randomly combined by cross-over and/or mutation, with the properties of SOLs with high fitness values more likely to pass on to the next generation. Each generation of SOLs is thus likely to attain overall higher fitness values, and the process is repeated until an optimal solution is reached. This could be when no or little improvement is gained in the fitness values between populations. A large problem space, such as all possible SOL configurations, can therefore be searched (non-exhaustively) in a reasonable length of time, though locally-optimal solutions may result. Starting with a diverse initial population and repeating the process for different initial populations should help guard against this.

A series of results using GAs to optimise SOLs are presented by Liu et al including a binary amplitude SOL (using the vectorial angular spectrum method) [99], and binary amplitude or phase SOLs giving both a scalar treatment [100] and a vector treatment [101]. The amplitude and phase modulations are implemented by restricting the transmission function to take values in $\{0, 1\}$ or $\{-1, 1\}$, respectively, and an outline is also given for extending the method to multi-amplitude SOLs [101]. In all cases, the single objective of minimising spot size is used, while constraining the intensity in the FoV, $I_{\text{rel}}$. Where a vectorial treatment is used, the method is applicable to radially-, circularly- and linearly-polarised light. Simulations and experimental demonstrations using these SOLs in confocal scanning microscopy are also included.

This approach is modified to construct optical needles with a binary amplitude metasurface [102], though this is later refined to a technique based on Fresnel zone plates that no longer uses a GA [103].

Chen et al use a GA to optimise binary amplitude (transmittance values $0$ or $1$), continuous amplitude (transmittance values between $0$ and $1$), binary phase (modulation values $0$ or $\pi$) and continuous phase (modulation values between $0$ and $2\pi$) SOLs [104]. A vector treatment is used to consider circularly- and radially-polarised light. They minimise the FWHM, imposing constraints on the FoV, $I_{\text{rel}}$, and spot intensity. Several examples are given and compared, with continuous phase SOLs demonstrating improved performance over amplitude SOLs. The authors also note that using radially-polarised light improves the spot size.

All four combinations of continuous and binary SOL with both amplitude and phase modulation are optimised by Wen et al [105]. This is done for a 1D SOL in the context of a manufacturing method demonstrated by He et al [106], where a double-layer hole array is used to simultaneously control amplitude (transmittance values between $0$ and $0.3$) and phase (modulation values between $0$ and $\pi$). For a specified FWHM, focal length and physical parameters, the objective of the GA is to find a SOL design with an intensity profile that matches these values. The authors find advantages to using a continuous amplitude and phase SOL, though include full 2D simulations for a continuous amplitude and binary phase SOL based on the double-layer hole array.

Two further methods for manufacturing 1D SOLs are given by Chen et al [107, 108]. The first enables continuous amplitude transmittance between $2\%$ and $98\%$, and the second combined binary amplitude transmittance (taking values $0$ or $1$) and phase modulation (with values $0$ or $\pi$). These are optimised by a similar approach to that of Wen et al [105] and demonstrated experimentally, both achieving a large FoV. The latter is also extended to 2D [109], analysed for circularly-polarised light and shown to achieve a long focal length.

Wan et al use a GA to optimise a binary phase SOL, where the value of the phase shift forms part of the optimisation problem [110]. Building on a GA formulation by Menon et al [111], the fitness function combines the energy in the spot, energy in the sidebands and the FWHM. These elements are delicately weighted (with a negative weight for the spot energy so that small values of the fitness function are preferred), and a step-wise optimisation regime is advised. This involves including only the energy considerations initially, introducing the FWHM and increasing the number of zones in later stages. A finite-difference time-domain simulation of the optimised SOL is also included.

Genetic algorithms are applied in a number of further scenarios. Diao et al design a binary amplitude SOL for realising an optical needle [112]. A single objective is optimised, comprising a carefully weighted sum of three terms. These three terms aim to maintain a uniform intensity of the optical needle, maximise the energy in the optical needle and maintain a constant FWHM. Additional constraints impose pre-specified parameters including the FWHM, depth of focus and focal length.

A similar approach is used to design a binary amplitude SOL for realising multi-focal superoscillatory spots [113]. Three weighted objectives, aimed at concentrating the energy uniformly in the spots, are combined into a single-objective optimisation problem. Additional constraints impose pre-specified parameters including FWHM and $I_{\text{rel}}$. 

This approach is then modified to implement the constraints as penalty functions rather than constraints, resulting in a single-objective unconstrained optimisation problem with improved convergence [114]. Binary phase SOLs are designed, with the spot size now also part of the objective. Multi-focal and achromatic superoscillatory spots, as well as a SOL aimed at tackling dispersion, are realised experimentally.

5. From theory to application

Many of the methods for constructing superoscillations reviewed above have been realised in experiments. However, the transfer of theory to experiment is never straightforward. There are numerous practical issues to be resolved beyond the issues arising from the well-known theoretical trade-offs. These include dealing with axial (and other) misalignment, imperfect manufacturing, noise and discretisation of continuous functions, to name but a few. Where superoscillations have been demonstrated experimentally, these issues have been overcome, and many authors include calculations and/or simulations to demonstrate the robustness of their designed superoscillations.

Mathematical and numerical methods must also be intrinsically robust to be of general use. This includes considerations such as induced instability, bounds on truncation errors and ensuring convergence where appropriate. Again, authors generally address such issues alongside the development of any new methods.

This must be taken seriously in any pursuit of applying these methods, and is sufficiently important that some papers are published entirely based on such considerations [115–118]. They are, however, beyond the scope of this review, and we refer the reader to individual publications of interest for further detail.

6. Applications of optimised superoscillations

The generality of the superoscillation concept—the fact that it allows unusually rapid oscillations in any band-limited wave system—gives it a huge range of potential applications, with some unexpected results. Systems of interfering waves, and in particular band-limited waves, are not just found in optics, but in all areas of science and technology. We briefly review some application areas below, to give an idea of the range of problems to which the mathematical techniques above have been applied. We aim to give a flavour, rather than an exhaustive list of applications of superoscillations, for which we refer the reader to other recent reviews [23, 24, 26].

6.1. Imaging

One of the more obvious and widely exploited application areas is super-resolution imaging, which is an enabling technology of immense importance across the physical and life sciences, as well as a field of research in its own right (for a recent review of super-resolution imaging, see [119]). Indeed, Berry noted the power of superoscillations for imaging as early as 1994, but it was 2007 before experimentalists described complex optical fields as superoscillations [10], with the same authors proposing an imaging technology soon afterwards [12]. In fact, Di Francia had already noted the application of these ideas to imaging in 1952 [9], although he did not use the terminology of superoscillations, and the body of work on PFs (see section 3) expanded on his idea and optimised the technology.

Microscopy is ideally suited to superoscillatory imaging, due to the degree of control of the whole optical system, and the high-power illumination sources available, compared to detector sensitivities: a standard 100 mW visible laser produces $10^{17}$ photons per second, and a typical microscopy camera is capable of producing high contrast images with tens of photons per pixel. Similarly, confocal scanning systems rarely use more than a milliwatt of laser power, leaving several orders of magnitude of power to compensate for the inevitable loss of energy into the sidebands. Confocal scanning microscopes also allow the sidebands to be filtered out using pinholes, though it has also been shown 

that widefield superoscillatory microscopes can achieve optical super-resolution. In 2013, Rogers et al. [88] published the first work using SOLs for imaging, and since then there have been many more advances in lens design (see section 4), lens fabrication [114, 120] and imaging technology [69, 121–123]. In a different approach, multi-lobe superoscillations have also been applied to structured illumination microscopy [49]. For a recent review of superoscillatory imaging see reference [23].

Although SOLs work well in transmission geometries, Nagarajan et al. [124] recently studied them for reflection confocal nanoscopy, finding the double pass through the SOL to degrade imaging performance. For this reason PF approaches, where the PF can be applied to only the focussing path, have been preferred for reflection microscopy [73, 125, 126].

In general SOLs suffer from chromatic aberration: they focus different colours of light into different focal spots, and often in an unpredictable pattern. More recently however, achromatic SOLs have been demonstrated for super-resolution focussing [91, 127] and imaging [128]. In applications such as
spectrometry, however, chromatic dependence is an advantage, and SOLs can be exploited for compact spectral imaging [114].

Telescopes, while more sensitive to loss than microscopes, have the same resolution challenges and several works have shown that superscaryllatory PFs can also enhance telescope resolution [129, 130].

6.2. Superoscillatory optical needles

Standard methods of creating superoscillatory spots tend to produce spots that do not persist in the axial (propagation) direction. This can be seen as an advantage in imaging applications (it can give high axial resolution) or a disadvantage (it gives short depth of focus). To extend the depth of focus, one approach is to create an optical needle: a superoscillatory focus that extends for many wavelengths in the propagation direction. These have been created with SOLs [89, 112, 131], and PFs [132–135], or by techniques involving the superposition of Bessel beams [75, 96, 136]. Roy et al demonstrated that these lenses could be used for subdiffraction-limited imaging [137]. Further work has designed SOLs specifically for use in the next generation of hard–disk drives, where tight focussing of UV light, but with a long depth of focus, is a major technological challenge [90, 138]. These optical needles could also be applied to nanofabrication with light, including optical micro- and nano-drilling, particularly where high length-to-diameter-ratio holes are needed [139], and are particularly well suited to the (very common) paradigm of optically-written planar fabrication, where a long depth of focus is a major technological advantage.

6.3. Optical lattices and trapping

Another common use of focussed light beams is for the trapping and manipulation of micro- and nano-particles (see [140] for a recent review of the field). As such it is a natural field for application of superoscillations, as noted by many authors [26, 70]. Axially invariant [141] or axially multifocal beams [114] hold particular promise and over the last few years, both optical [142–144] and acoustic [145] superoscillatory traps have been demonstrated.

Trapping multiple particles (using optical lattices) is a natural field where superoscillatory patterns (with multiple spots in the focal plane, rather than a single focal spot) are of particular use. These types of field have been studied from the early demonstrations of superoscillations, where quasi-crystalline arrays of spots were investigated [10, 12]. Shortly after superoscillatory lenses became common, lens arrays (creating spot arrays) were also designed and demonstrated [146]. More recently, the Talbot effect has been reported in superoscillatory spot patterns [147]. While multilobe superoscillations have already been used for structured illumination microscopy [49], it would be interesting to investigate the use of superoscillatory spot-arrays for multifocal structured illumination microscopy [148].

6.4. Optical superoscillations in the time domain

As well as the well-known problem of focussing light into small spots, the creation of short temporal pulses is a key challenge in both optics and signal processing. Conventionally, the minimum pulse length is limited by the frequency and bandwidth of the optical field: analogous to the spot size being limited by the range of spatial frequencies available. Superoscillatory pulses, therefore, can break this limit and produce rapid temporal oscillations [59] and temporally compressed pulses [29, 149]. These temporal superoscillations can be used to transmit high-frequency signals through absorbing media [59] or low pass filters [150], and have been proposed to increase the range resolution of a radar system [151, 152] or to allow super-narrow frequency conversion [48].

6.5. Suboscillations

As well as superoscillations, which oscillate faster than their highest Fourier component, suboscillations, which oscillate slower than their lowest Fourier component have been proposed and studied. Eliezer and Bahabad developed a mathematical treatment and demonstrated optically that they can cause light to defocus unexpectedly rapidly [153]. Further work has investigated the mathematics [154] of suboscillations and shown that arbitrarily shaped suboscillations can be designed [155].

6.6. Non-optical

Although many of the applications of superoscillations demonstrated so far have been in optics, it should be remembered that they were originally described in quantum mechanics [2] where research in superoscillations continues to push boundaries. For example, Aharonov et al have proved a surprising longevity of time-evolving superoscillatory wavepackets [156, 157], and there is continual potential for the applicability of such new results from quantum mechanics to other application areas.

Many further applications outside optics have also been explored. Superoscillations have been applied to terahertz focussing [158, 159], used to create super-resolved electron beams [160] and used to create
subwavelength acoustic foci [145, 161, 162]. In fact, the work of di Francia [9] was inspired by previous work on radio-frequency super-gain (or super-directive) antennas. These use the same mathematics as superoscillations but in reciprocal spaces. Optical superoscillations have limited spatial frequencies (or equivalently angles of plane waves) and yet produce unexpectedly small spatial features. Super-gain antennas use spatially-limited antennas to produce beams with unexpectedly small angular distributions [60, 61].

From the point of view of applications however, there is one crucial distinction: in super-gain antennas, the (angular) sidebands can be pushed beyond an angle of 90°, where they enter the evanescent region and are not transmitted. In this case their effect is seen as a reactive loss (the antenna has to be supplied with more energy than is transmitted) but the sidebands do not propagate and hence are not seen in the far-field. In superoscillations, on the other hand, the sidebands remain real and are seen in the same space as the superoscillatory feature.

Data compression is another obvious area of application for superoscillations: if we can (locally) increase the effective bandwidth, then surely we can send more information through a band-limited transmission channel. It is important to note, however, that the overall channel capacity is actually both bandwidth limited and noise-limited: that is, the bandwidth limit is valid only for a particular noise level. If we consider a noisy channel, the weaker superoscillatory section of the pulse becomes more difficult to faithfully detect, requiring a lower-noise channel for error-free transmission. However, it is well-known that lowering the noise floor also increases the channel capacity [163, 164] by the same amount as the superoscillatory compression would achieve. The overall effect is that ‘compressing’ data by encoding into a superoscillatory signal does not increase the overall channel capacity [36, 165], as might be naively expected, although it provides a way of compressing data differently, which may have subtle benefits in some scenarios.

Broadening the scope even more, superoscillations have been proposed as a mechanism for heating in the solar corona [166, 167], and for super-resolution in ground penetrating radar for landmine detection [34, 168].

7. Concluding remarks and future outlook

This review is a tool for anyone planning to apply superoscillations to a new application, or to extend the boundaries in an application where superoscillations have already been used. It lays out the available mathematical methods, their considerations and capabilities, so that the one that is best suited to a particular purpose can be chosen. Whether it is a new application that requires a large FoV with multiple superoscillations, the design of a PF that is easier to manufacture, or the need for a spot that persists on propagation, there are a multitude of available design methods at our disposal.

We have minimum-energy and minimum-energy-yield solutions, we have solutions that analytically optimise some combination of spot size, spot intensity and FoV. Other solutions exploit the properties of Bessel beams to achieve large depth of focus, or the cosine function for a smooth and regular result. Numerical methods simultaneously tackle multiple objectives, and include several superoscillatory spots and wavelengths. It is not surprising that we see a trend over time from analytic methods considering relatively simple cases, using methods from linear algebra and variational techniques, to numerical methods tackling increasingly complex scenarios with genetic algorithms.

Not only do these methods provide a range of options suited to different purposes, they also describe the inevitable trade-offs that must be faced in applying superoscillations. Each application, and each scenario, will be sensitive to different aspects of these trade-offs, and have different tolerances. Energy loss is not a problem if powerful lasers are used with only a fraction of the available energy actually needed. However, the relative intensity of the sidebands may pose a significant challenge. The shape of the superoscillations may be crucial in some applications, whereas in others it can be relaxed. This review also outlines the capabilities of each method to tackle such trade-offs, a consideration which will be equally important in choosing the method best suited to a particular purpose.

Despite the numerous different technological methods for realising superoscillations, the mathematical design methods can be categorised into just three different approaches: the direct approach of designing a superoscillatory function, or specifically setting out to design a PF or a SOL. The theoretical capabilities of the direct approach are limited only by our current mathematical understanding, while starting with a particular PF or SOL in mind inevitably imposes restrictions. A PF or SOL may, however, be the practical solution. The mathematics behind PFs is relatively straightforward, with the use of a lens enabling far-field approximations to be employed. Designing SOLs tends to lead to difficult inverse problems that are more difficult to solve. An improved theoretical understanding of the mechanisms behind SOLs, and further insight into inverse problems, will aid development in the area of SOL design.

Alongside this development of mathematical methods, the technology for realising superoscillatory functions has also seen major breakthroughs. Recent metalenses can simultaneously and (almost)
continuously modulate both amplitude and phase, making the direct approach of designing superoscillatory functions even more applicable to the search for the optimal SOL, while arbitrary pupil filters can be realised with SLMs.

As well as laying out the available methods for designing superoscillations, this review also gives a picture of the field of applied mathematics within the superoscillations literature, highlighting some areas for development. Further theoretical analysis will enable superoscillations to be more fully exploited in applications. This includes studying superoscillatory functions further in their own right, to really probe our understanding of their properties, as well as optimising the features of superoscillations that are relevant to applications. While the trade-off between the number of superoscillations and size of superoscillations with total energy is understood in 1D, this has not been fully described in 2D or 3D. Furthermore, such relationships between measures that may be of more concern in applications, such as FoV and relative intensity, could be better characterised. Fundamental limits on sets of best possible compromises imposed by the various trade-offs are also not known. With new application areas, new considerations and trade-offs are also likely to emerge.

With sophisticated numerical methods now available, alongside widely-available processing power, it is tempting to be led down that route, and no doubt much future development will result from such pursuits. However, it is important not to forget the rich body of theoretical results and understanding that has built up over the past decades, and to ensure that this is properly utilised in future developments, which may include analytic and semi-analytic approaches.

It is an exciting time in the field of superoscillations, as early theoretical analysis has evolved to making a growing impact on an ever-increasing number of real-world applications. There is great scope for both theoretical growth to further our understanding of this remarkable effect, as well as technological break-through to further push boundaries.

**ORCID iDs**

K S Rogers  [https://orcid.org/0000-0001-9641-6032](https://orcid.org/0000-0001-9641-6032)

E T F Rogers  [https://orcid.org/0000-0001-6166-4087](https://orcid.org/0000-0001-6166-4087)

**References**


[28] Eliezer Y and Bahabad A 2016 Super-oscillating airy pattern ACS Photonics 3 1053–9
[48] Remez R and Arie A 2015 Super-narrow frequency conversion Optica 2 472
[56] Huang F M and Zehluev N I 2009 Super-resolution without evanescent waves Nano Lett. 9 1249–54
[61] Dolph C L 1946 A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level Proc. IRE 34 335–48
[66] Wong A M H and Eleftheriades G V 2017 Broadband superoscillation brings a wave into perfect three-dimensional focus Phys. Rev. B 95 075148
[120] Halim N et al 2018 Large-scale high-numerical-aperture super-oscillatory lens fabricated by direct laser writing lithography RSC Adv. 8 20117–23
[121] Yin F, Chen G, Chen W, Qiao W and Guan J 2020 Superresolution quantitative imaging based on superoscillatory field Opt. Express 28 7707
[123] Wenli Li, Pei H, Yuan W and Yiting Y 2020 Efficiency-enhanced and sidelobe-suppressed super-oscillatory lenses for sub-diffraction-limit fluorescence imaging with ultralong working distance Nanoscale 12 7063–71
[127] Zhu Li et al 2018 Achromatic broadband super-resolution imaging by superoscillatory metasurface Laser Photon. Rev. 12 1–6
[140] Liu J and Zhuyuan Li 2018 Controlled mechanical motions of microparticles in optical tweezers Micromachines 9 232
[152] Berry M V 2016 Representing Superoscillations and Narrow Gaussians with Elementary Functions Milan J. Math. 84 217–30
[153] Eliezer Y and Bahabad A 2017 Super defocusing of light by optical sub-oscillations Optica 4 440
[158] Legaria S, Pacheco-penha V and Beruete M 2018 Super-Oscillatory Metalens at Terahertz for Enhanced Focusing with Reduced Side Lobes Photonics 5 56
[159] Ruan D et al 2018 Realizing a terahertz far-field sub-diffraction optical needle with sub-wavelength concentric ring structure array Appl. Opt. 57 7905


