Structured Illumination Microscopy

7

Nicholas Hall and Ian Dobbie

Contents

7.1 What You Should Already Know 174
7.2 Fourier Decomposition of Images 174
7.3 Optics in Reciprocal Space 179
7.4 Principles of SIM 184
7.4.1 Reconstructing a SIM Image 190
7.5 Strengths and Limitations of SIM 192
References 192

What You Will Learn in This Chapter

The advent of super-resolution techniques has fundamentally changed how biology is done and what biological questions it is possible to ask and answer now that we can image beyond the resolution limit. There are a number of super-resolution techniques, each with their own strengths and preferred use cases. Structured illumination microscopy (SIM) relies on spatially structured illumination light to encode super-resolution information in a resolution limited image. A number of these resolution limited images are used together to extract the super-resolution information, producing a reconstructed image with twice the resolution of the original images. The number of images required to obtain a super-resolution image is considerably less than most other super-resolution techniques, making SIM ideally suited to imaging dynamic biological processes. Herein we will cover how to think of images as information in Fourier space (including an introduction to Fourier

N. Hall (✉)
TCAD, Synopsys Inc, Glasgow, UK
e-mail: nickhall@synopsys.com

I. Dobbie
Integrated Imaging Center, Department of Biology, Johns Hopkins University, Baltimore, MD, USA
e-mail: ian.dobbie@jhu.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

V. Nechyporuk-Zloy (ed.), *Principles of Light Microscopy: From Basic to Advanced*, https://doi.org/10.1007/978-3-031-04477-9_7

mathematics sufficient to understand this), the mathematics behind SIM and how images are reconstructed, and the strengths and limitations of SIM.

7.1 What You Should Already Know

There are a few of concepts that the reader is assumed to be familiar with and/or have been covered in previous chapters. These are:

- Geometric optics leading to the rise of the diffraction limit
- Fluorescence microscopy
- An understanding that functions can be represented by Fourier decomposition

These will be explained briefly when and where they are relevant. However, the full scope and complexity will not be covered. Readers who are unfamiliar with any of these are advised to pause here and familiarise themselves. Some suggested sources are provided in the *Further Reading* section at the end of the chapter.

7.2 Fourier Decomposition of Images

In order to understand SIM, we must first take a step back and ensure we understand a more fundamental concept, Fourier analysis. What follows will not be a series of extensive mathematical derivations. These can be found in previous publications. Rather specific formulae will be stated when appropriate to demonstrate particular important details.

Readers should already be familiar with the concept of Fourier decomposition; describing a function as an infinite series of frequency components, also referred to as reciprocal components, called a Fourier series. Typically, this explanation starts with time-varying functions since most people have an intuitive understanding of a single time-varying function being composed of a number of frequency components. For example, it is common knowledge that a single musical chord is composed of a number of different notes (i.e. frequency components) played simultaneously. However, there is nothing unique about time-varying functions. If the function is spatially varying rather than time-varying, all of the mathematics remain unchanged. In this case, the Fourier components would be described by reciprocal quantities that we call *spatial frequencies*.

More generally, we can describe a reciprocal variable for the Fourier series to correspond to the original variable in the function. Likewise, in the same way, we describe a vector space for the original variable or variables—such as the 3-dimensional vector space described by the *xyz* Cartesian vectors—we can describe a vector space for the reciprocal variables which we call reciprocal space. Therefore, we can say:

Any function can be described by an infinite Fourier series of reciprocal variables.

There is a mathematical operation to convert from the original vector space to the corresponding reciprocal space known as the Fourier transform which is written as [1]

$$\widetilde{\mathbf{F}}(\boldsymbol{k}_x) = \int_{-\infty}^{\infty} f(\boldsymbol{x}) e^{-i\boldsymbol{k}_x \boldsymbol{x}} d\boldsymbol{x}, \tag{7.1}$$

where $\boldsymbol{x}$ and $\boldsymbol{k}_x$ are the original and reciprocal variables, respectively. This yields a continuous function, $\widetilde{\mathbf{F}}(\boldsymbol{k}_x)$ described in terms of the reciprocal variable, $\boldsymbol{k}_x$. In general $\widetilde{\mathbf{F}}$ is used to denote the Fourier transform of f. Likewise, the operation to move from reciprocal space back to the original vector space, the inverse Fourier transform, is written as

$$f(\boldsymbol{x}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widetilde{\mathbf{F}}(\boldsymbol{k}_x) e^{i\boldsymbol{k}_x \boldsymbol{x}} d\boldsymbol{k}_x, \tag{7.2}$$

Figure 7.1 shows the top hat function and its Fourier transform, a sinc function. Notice that the Fourier transform has positive and negative frequency components and it is symmetrical around $\omega = 0$. This symmetry is the result of a more general property of Fourier transforms of real functions which is that:

$$\widetilde{\mathbf{F}}(\boldsymbol{k}_x) = \overline{\widetilde{\mathbf{F}}(\boldsymbol{k}_x)}, \tag{7.3}$$

where $\overline{\widetilde{\mathbf{F}}(\boldsymbol{k}_x)}$ is the so-called complex conjugate of $\widetilde{\mathbf{F}}(\boldsymbol{k}_x)$. A complex conjugate of a complex number $a + ib$ is $a - ib$. Essentially the value of the Fourier transform at some value of $\boldsymbol{k}_x$ is the complex conjugate of the value of the Fourier transform at $-\boldsymbol{k}_x$. Now, the Fourier transform of a top hat function symmetric around 0 has no imaginary component (i.e. $b = 0$ in $a + ib$) and so this property becomes $\widetilde{\mathbf{F}}(\boldsymbol{k}_x) = \widetilde{\mathbf{F}}(-\boldsymbol{k}_x)$, and is symmetric around $\omega = 0$.

Although up to this point, we have dealt with one-dimensional functions, there is nothing in the mathematical construction of the Fourier transforms which requires this. Take, for example, the Mona Lisa shown in Fig. 7.2a. Although it might not be immediately obvious, there exists a mathematical function which describes the intensity variations in this image. We can, therefore, consider the function which describes the Mona Lisa as the visualisation of a 2D function in $(\boldsymbol{x}, \boldsymbol{y})$. As such, we can also perform a Fourier transform on this function to acquire Fig. 7.2b, which is a function in $(\boldsymbol{k}_x, \boldsymbol{k}_y)$ where $\boldsymbol{k}_x$ and $\boldsymbol{k}_y$ are the spatial frequencies which are reciprocal variables to the x and y variables, respectively. Once again, these are equivalent representations of the same information. We only use different variables to describe said information.

It is worth noting that the Fourier space representation is equivalent to the real space representation (i.e. the actual function), just described in terms of reciprocal variables. A mathematical theorem, the derivation of which is not presented here,

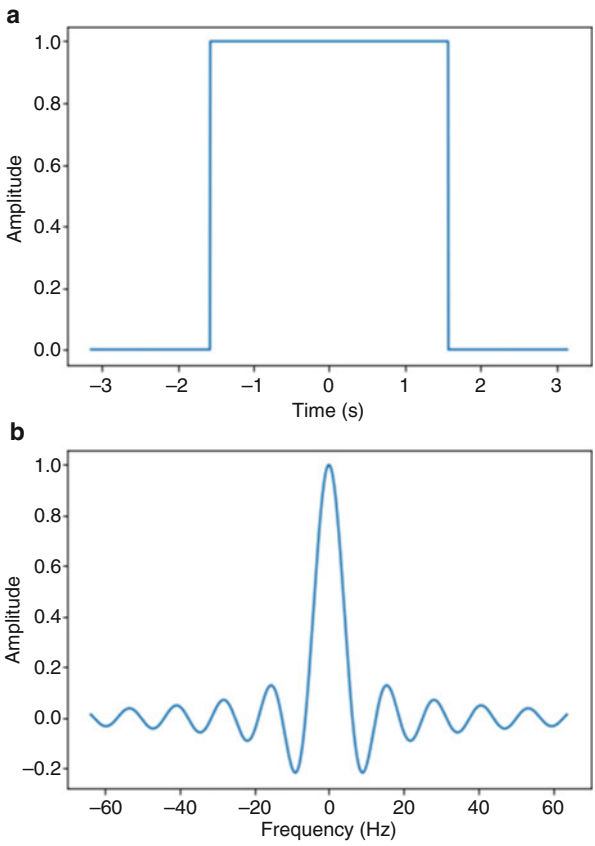

Fig. 7.1 (**a**) A continuous top hat function centred at the origin, $f(x)$. (**b**) The continuous Fourier transform of $f(x)$, $\widetilde{\mathbf{F}}(\boldsymbol{k}_x)$

called Parseval's theorem states that the integral of the square of a function is equal to the integral of the square of its Fourier transform. Now, when the function in question describes amplitude—in the case of light, the amplitude of the electromagnetic field—then the square of the function is its intensity. Therefore, the integral of the square of the function f is the total energy of the signal or image the function describes. Consequently, the integral of the square of both the real space function and its Fourier transform is the total energy of the signal or image the function describes. This information allows us to conclude that the Fourier space representation is equivalent to the real space representation, just described in terms of reciprocal variables.

We should also note that the previous discussion refers to continuous functions, that have a known value at all possible positions in space, and that these need infinite Fourier components to be faithfully reproduced. However, a modern microscope image is represented by pixels in a camera image, which is only defined at the centre positions of the pixels. This type of function can be completely defined by a finite Fourier series with the same number of components as the positions in the original discrete function.

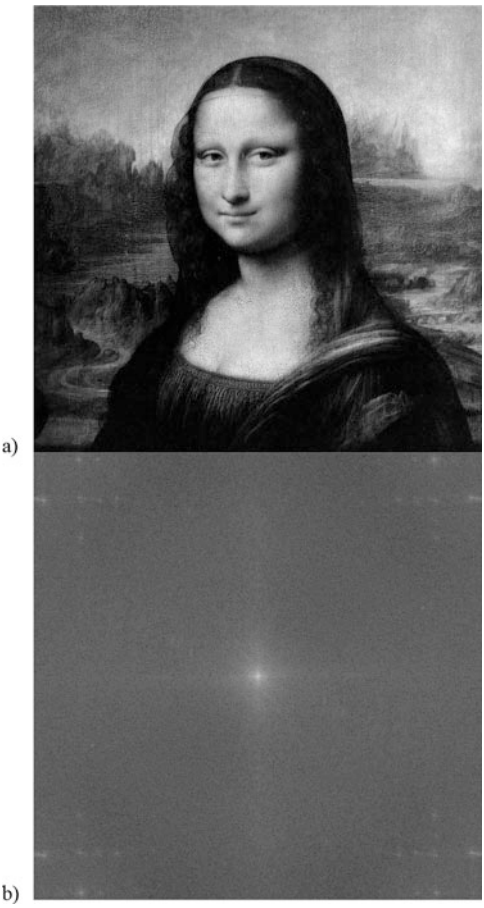

Fig. 7.2 (**a**) An greyscale image of the Mona Lisa. (**b**) The 2D Fourier transform of the image

Before we proceed, we also need one last mathematical concept; *convolution*. The convolution operation is defined as

$$f(x) \circledast g(x) = \int_{-\infty}^{\infty} f(\tau) g(x - \tau) d\tau = \int_{-\infty}^{\infty} f(x - \tau) g(\tau) d\tau, \tag{7.4}$$

where $\circledast$ is the convolution operator. The resulting function expresses how the shape of one function modifies the other. It may not be intuitively apparent what this means, which is fine as a complete understanding of convolution is not required at this stage. There are two relevant properties of convolution which are required going forward.

The first is what happens when some function, $f(x)$, is convolved with a Dirac delta function, $\delta(x - a)$. The Dirac delta function is a real-valued function which is 0 everywhere except at the position a, where it is 1. In other words, it is an infinitely

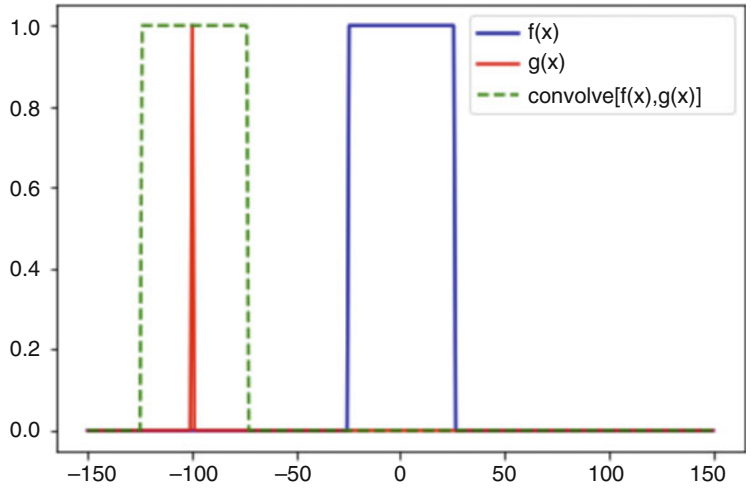

Fig. 7.3 The convolution of a top hat function centred at the origin, $f(x)$, with a Dirac delta function centred on -100, $g(x) = \delta(x + 100)$. The resulting convolution is a top hat function centred on -100, i.e. $f(x + 100)$

narrow peak at a particular position, a, and 0 everywhere else. When some $f(x)$ is convolved with the Dirac delta function, we obtain

$$f(x) \circledast g(x) = \int_{-\infty}^{\infty} f(\tau)\delta(\tau - a)d\tau = f(x - \tau), \tag{7.5}$$

In other words:

The convolution of a function $f(x)$ with a Dirac delta function results in $f(x)$ being shifted to be centred at the position of the Dirac delta function.

Figure 7.3 show the convolution of a top hat function centred at the origin, $f(x)$ with a Dirac delta function centred on -100, $g(x) = \delta(x + 100)$. As we can see, function obtained as a result of this convolution is identical to the top hat function, except it is centred on $x = -100$ instead of $x = 0$.

The other property of convolutions is the *convolution theorem*. Say we once again have the convolution of two functions, $f(x)$ and $g(x)$. The convolution theorem states that:

$$\mathcal{F}\{\, f(x) \circledast g(x)\} = \widetilde{\mathbf{F}}(\boldsymbol{k}_x) \cdot \widetilde{\mathbf{G}}(\boldsymbol{k}_x), \tag{7.6}$$

where $\mathcal{F}$ denotes performing a Fourier transform (otherwise known as the Fourier operator), $\cdot$ is the dot product (i.e. point-wise multiplication), and $\widetilde{\mathbf{F}}(\boldsymbol{k}_x)$ and $\widetilde{\mathbf{G}}(\boldsymbol{k}_x)$

are the Fourier transforms of the functions $f(x)$ and $g(x)$, respectively, as before. What the convolution is essentially stating is:

The Fourier transform of a convolution of two functions is equivalent to the product of their Fourier transforms

From Eq. (7.6), we can obtain the corollary:

$$\mathcal{F}\left\{\widetilde{\mathbf{F}}(\boldsymbol{k}_x)\circledast\widetilde{\mathbf{G}}(\boldsymbol{k}_x)\right\} = f(x)\cdot g(x). \tag{7.7}$$

7.3 Optics in Reciprocal Space

Most readers should be familiar with the concept of the resolution limit in microscopy. The typical formulation of the resolution limit is the minimum separation at which two point objects can be meaningfully separated as independent objects. The often-quoted value for this resolution limit from the Rayleigh Criterion is

$$r_l \approx \frac{1.22\lambda}{2NA}, \quad r_a \approx \frac{2\lambda n}{\mathrm{NA}^2}, \tag{7.8}$$

where r_l and r_a are the lateral and axial resolution limits, respectively, λ is the wavelength of light, NA is the numerical aperture of the imaging lens, and n is the refractive index of the media [2–5]. These formulations require several approximations and simplifications, but they are nonetheless useful quantities as the theoretical, ideal resolution limit of a microscopy system.

Rather than exploring the derivation of these diffraction limits from a geometric perspective, it is useful for us to instead discuss their origins using the notions of reciprocal space we have established previously. This will allow us to more intuitively understand the principles of SIM. First, consider a single slit with monochromatic light shining through it as shown in Fig. 7.4 which should be familiar to readers. Consider a point which is vertically displaced y from the centre of the slit with width d and a distance L from the slit. There is a path difference, shown in red in Fig. 7.4, between the top and bottom of the slit determined by $d \sin\theta'$.

In the extreme case of a maxima in the diffraction pattern, this path difference must be an integer multiple of the wavelength of light, λ, i.e. $d \sin\theta' = m\lambda$, where m is some integer. Using the simple geometry $\tan\theta = \frac{y}{L}$, the approximation that $\theta \approx \theta'$ and the paraxial approximation that $\sin\theta \approx \tan\theta \approx \theta$ then the maxima, or anti-nodal, intensity positions are given by

$$y = \frac{m\lambda L}{d}, \tag{7.9}$$

Similarly, the minima, or nodal, intensity positions are given by

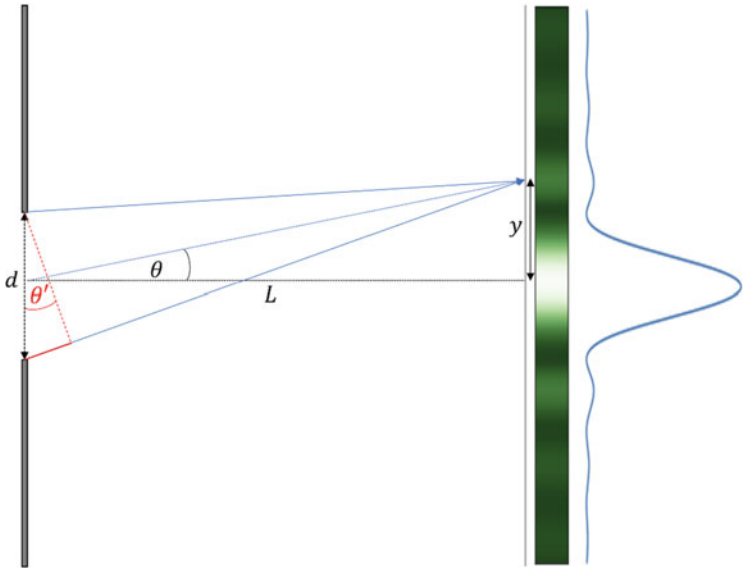

Fig. 7.4 Diagram of a single slit diffraction pattern

$$y = \frac{n\lambda L}{d}, \tag{7.10}$$

composed of binary minima and maxima, as the intensity line profile to the left of Fig. 7.4 shows. In fact, the intensity of the diffraction pattern, λ, is directly proportional to $\text{sinc}(y)^2$.

This follows if we consider a situation we have already discussed in Sect. 7.2; the top hat function. Previously we considered the amplitude of a time-varying top hat signal pulse. As shown in Fig. 7.1, the Fourier transform of this time-dependent function is a sinc function. Now, consider the transmissivity profile of a single slit. It is 0 everywhere except the slit opening, essentially a top hat centred around the centre of the slit. The amplitude profile of the electromagnetic field of the light is therefore also a top hat. As previously established, the Fourier transform of a top hat function is a sinc function, but we observe the intensity of the diffraction pattern not its amplitude. Since intensity is directly proportional to amplitude squared, it should not be surprising that the diffraction pattern is a sinc^2 function.

Now, consider the setup shown in Fig.7.5a where we place a lens at L with a focal length, $f = \frac{L}{2}$. At L we would expect to recover an image of the Fourier transform of the slit. If we then place an identical lens (i.e. a $\times 1$ magnification setup) at $\frac{3}{2}L$, then at $2L$ we might expect to recover a perfect image of the single slit, as shown in Fig. 7.5b. However, as the inclusion of the word "perfect" in the previous sentence might imply, this is not what happens. Instead at $2L$ a blurred image is observed similar to Fig. 7.5c.

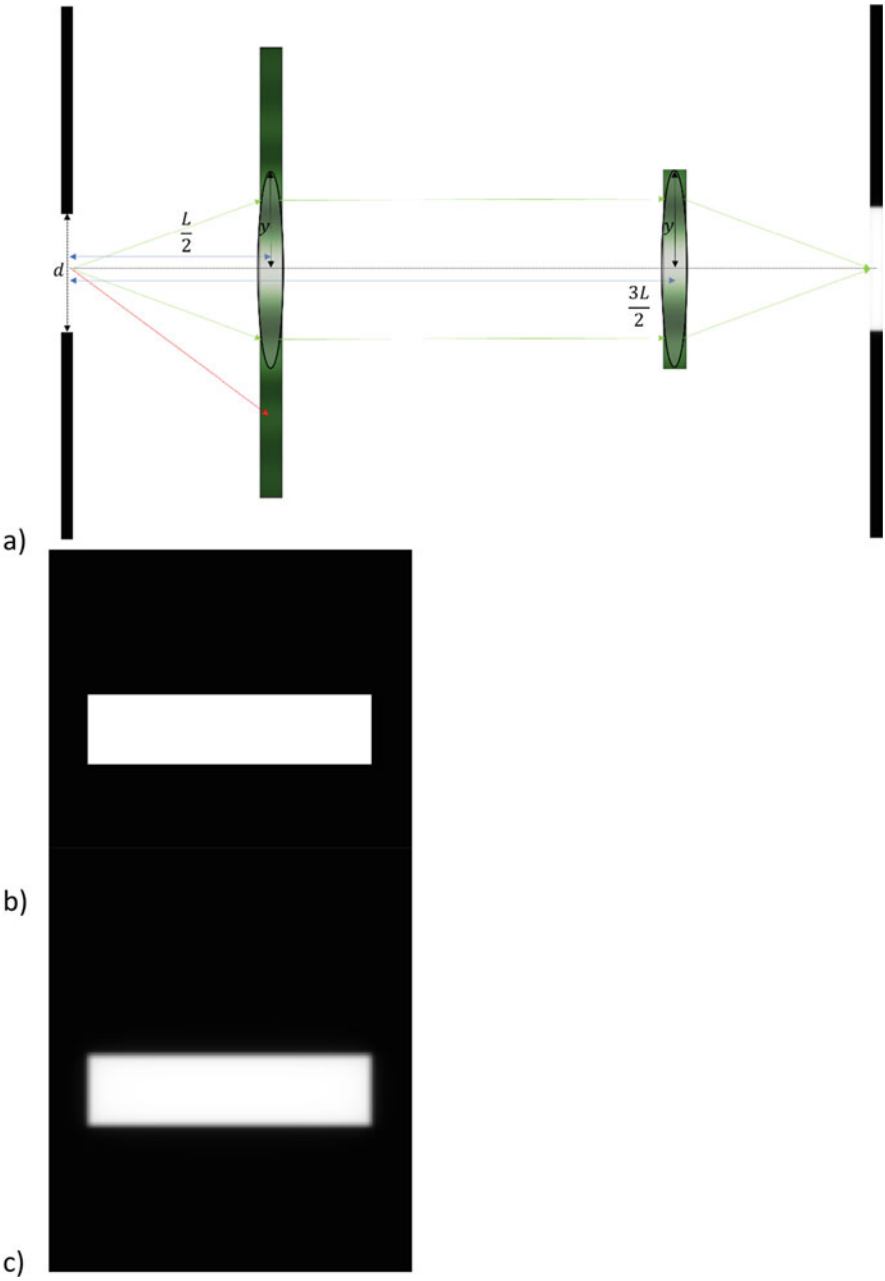

Fig. 7.5 A simple imaging system observing a single slit. (**a**) Diagram of a single slit imaging setup with a $\times 1$ magnification. Diffraction orders captured by the imaging setup are shown in green. Diffraction orders not captured by the imaging setup are shown in red. (**b**) Ideal image of the single slit. (**c**) Diffraction limited image of the ideal single slit

This occurs because a lens at L has a finite aperture. Therefore, only a finite portion of the diffraction pattern is collected by the imaging lens. In other words, only a finite range of spatial frequencies are collected by the imaging lens. This is shown in Fig.7.5a where certain diffraction orders are within the lens apertures and therefore propagate through the system, whereas others are not collected and therefore are missing when the final image is formed. When these spatial frequencies are used to construct an image at $2L$ by the second lens, the absence of the higher order spatial frequencies leads to an incomplete reconstruction of the original object, in this case a single slit. In effect, the imaging lens acts as a low-pass filter on the spatial frequencies. The spatial frequencies which are captured by the imaging system as said to be within the *observable region*.

In the lateral reciprocal plane, $\boldsymbol{k}_x\boldsymbol{k}_y$,—that is, the spatial frequency plane which is reciprocal to the lateral optical plane xy—the observable region has a radius ω_l as shown in Fig. 7.6a. In the simple example considered previously, this would be the radius of the imaging lens at L. Modern microscopy systems are considerably more complex than this and contain a number of lenses which collect the diffracted spatial frequencies. As such, we define a quantity, numerical aperture or NA, which describes the effective aperture of the entire imaging setup. The radius of the observable region in the $\boldsymbol{k}_x\boldsymbol{k}_y$ plane is defined as

$$\omega_l = \frac{2NA}{\lambda}, \tag{7.11}$$

which observant readers will note is the reciprocal of the Abbe diffraction limit [6]. The bounds of the observable region in $\boldsymbol{k}_z$ differs since the intensity spectrum in $\boldsymbol{k}_z$ is independent of the object nature and the point spread function (PSF) [5]. The bounds of the observable region in $\boldsymbol{k}_z$ can be shown to be

$$\boldsymbol{k}_z = \pm\frac{\left\|\boldsymbol{k}_{xy}\right\|}{2\lambda_k}\left(\omega_l - \left\|\boldsymbol{k}_{xy}\right\|\right), \tag{7.12}$$

for $\|\boldsymbol{k}_{xy}\| \leq \omega_l$, where $\|\boldsymbol{k}_{xy}\|$ is the length of the vector denoting the lateral spatial frequencies, $(\boldsymbol{k}_x, \boldsymbol{k}_y)$ and $\lambda_k = \frac{2\pi}{\lambda}$ [5]. Figure 7.6b shows the projection of the observable region onto the axial spatial frequency plane, $\boldsymbol{k}_x\boldsymbol{k}_z$. The maximum extension of this observable region, ω_a, is given by

$$\omega_a = \frac{\omega_l^2}{8\lambda_k}, \tag{7.13}$$

From Eq. (7.12), we can see that unlike the observable region for the lateral spatial frequencies, the observable region in $\boldsymbol{k}_z$ is bandpass limited at both high and low spatial frequencies. This results in the "missing cone" phenomenon which limits the axial spatial frequencies which are able to be collected, in turn limiting the axial resolution [7, 8]. This, coupled with the fact that for all practical cases $\omega_a \ll \omega_l$,

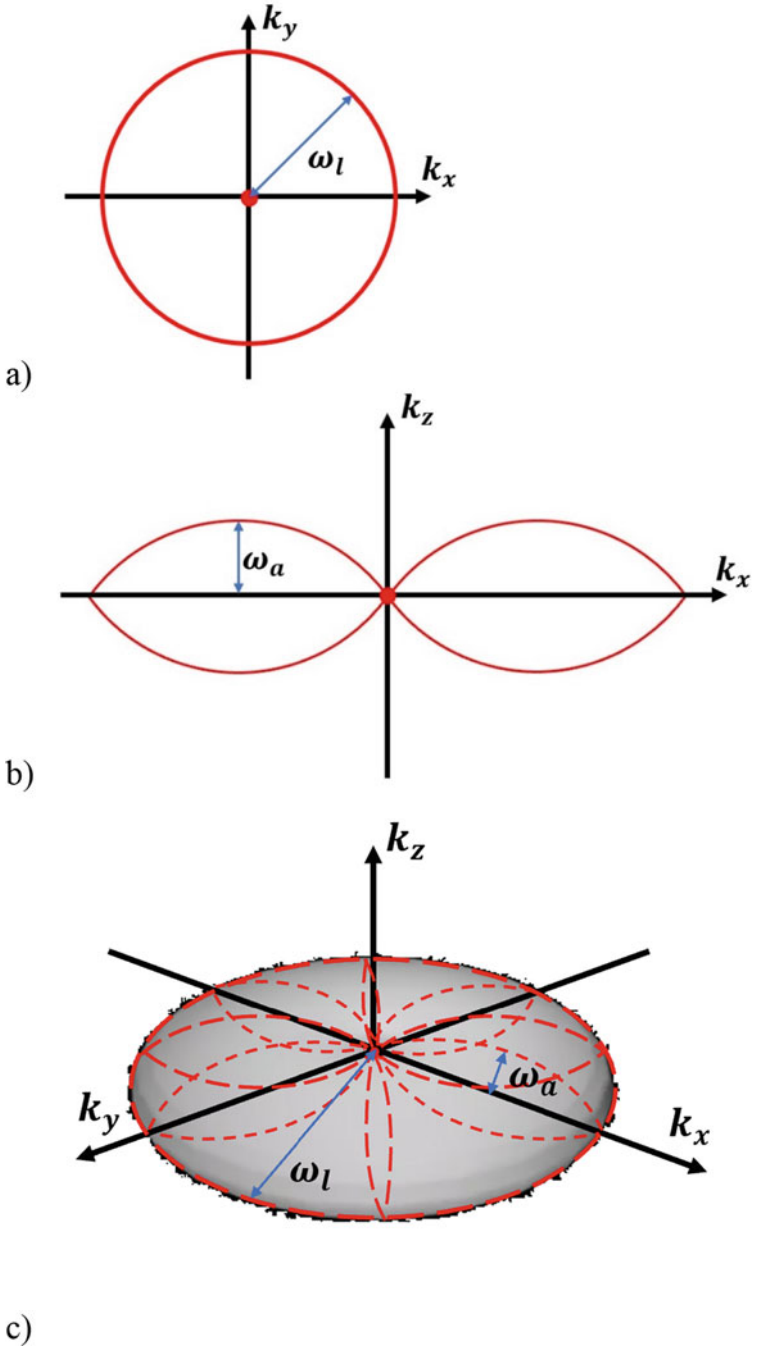

Fig. 7.6 Visualisation of the observable region of a conventional imaging system. (**a**) The projection of the observable region onto the $k_x k_y$ plane. (**b**) The projection of the observable region onto the $k_x k_y$ plane. (**c**) The full observable region for a conventional widefield microscope in reciprocal space as a 3D render

explains why the axial resolution is always worse than the lateral resolution of a system. Figure 7.6c shows the complete 3D observable region, obtained by rotating Fig. 7.6c around the $\boldsymbol{k}_z$ axis resulting in a torus-like 3D observable region which they can be described by

$$O_0(\boldsymbol{k}) = \begin{cases} 1, & if \left\|\boldsymbol{k}_{xy}\right\| < \omega_l \; and \; \|\boldsymbol{k}_z\| < \dfrac{\left\|\boldsymbol{k}_{xy}\right\|}{2\lambda_k}\left(\omega_l - \left\|\boldsymbol{k}_{xy}\right\|\right) \\ 0, & \text{otherwise} \end{cases}. \tag{7.14}$$

Only spatial frequencies within this observable region contribute to the observed image. If we consider an illumination wavelength of 561 nm and a numerical aperture of 0.7—typical values for a widefield air objective imaging system—we obtain $\omega_l = 2.50 \times 10^6\ \mathrm{m}^{-1}$. Since spatial frequencies have a corresponding peak-to-peak separation, for this imaging system we can see that objects which are laterally separated by less than $\frac{1}{\omega_l} = 400$ nm cannot be separated since no lateral spatial frequency with a peak-to-peak separation of less than 400 nm is captured by the imaging system. So produced images cannot contain objects with less than that lateral separation.

Individual objects which are below the resolution limit will still be observed, since a portion of their intensity is captured within the observable region and form an image. However, their apparent size will still be defined by the resolution limit, since the spatial frequencies required to construct an image with dimensions smaller than this are not available.

For a practical demonstration of the effects this observable region has on the image of a sample, take the ideal image presented in Fig. 7.7a. As shown in Fig. 7.7c, this image has a complete set of the spatial frequencies present in the underlying structure. Once a diffraction limit is imposed on the ideal image, Fig. 7.7b shows that although much of the large structures remain clearly visible, many of the fine structures such as spacing between individual leaves, small ripples of water, or patches of lichen on the rocks are no longer observable. In the power spectrum shown in Fig. 7.7d, the only spatial frequencies with non-zero values are those within the observable region.

7.4 Principles of SIM

What follows is not a rigorous derivation of the mathematical framework which underpins SIM imaging and reconstruction. Rather it is meant to provide readers with an intuitive understanding of the principles which are employed in SIM. For a more rigorous explanation of the mathematics, consult the papers references in the Further Reading section at the end of the chapter.

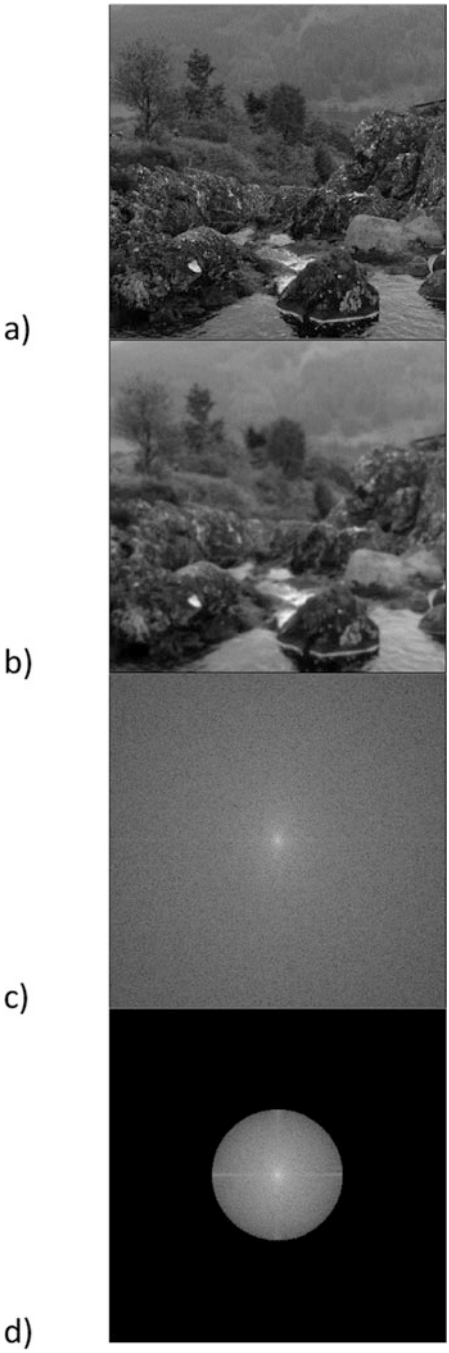

Fig. 7.7 The impact of a finite observable region on the resolution of an image. (**a**) an ideal image, (**b**) a resolution limited image, (**c**) the power spectrum of the ideal image. (**d**) The power spectrum of the resolution limited image. The only spatial frequencies which have any power in them are those within the observable region

This implementation of SIM is best understood through the moiré effect. Two patterns, such as those shown in Fig. 7.8a, b, which are simply two-dimensional positive sinusoids of the form:

$$f(x,y) = 1 + \cos\left(\tau_x x + \phi_x, \tau_y y + \phi_y\right), \tag{7.15}$$

where x and y are the lateral coordinates, and τ_x and τ_y are the frequencies of the sinusoid in the x and y axis, respectively, and are multiplicatively superimposed on one another will produce a beat pattern—moiré fringes—such as those shown in Fig. 7.8c. Using the convolution theorem, the Fourier transform of this superposition is equivalent to the convolution of the Fourier transforms of the two original patterns [9], as shown in Fig. 7.8d–f.

For SIM, one of the patterns is the underlying biological structure—or more specifically the spatial distribution of fluorophores—and the other is the spatially structured excitation illumination. The moiré fringes arising from the superposition of these two structures can be coarser than either of the original patterns, meaning that information arising from biological structures beyond the resolution limit of the microscope and a known illumination pattern can be observed. These moiré fringes contain information about these super-resolution structures and the super-resolution information can be extracted from the moiré fringes, effectively extending the observable region of a microscope beyond the diffraction limit [10].

In order to provide an intuitive understanding of SIM, let us look at a simulated example of 2D SIM. Here we will return to the ideal image from Fig. 7.7a. As we have already seen, only a portion of the spatial frequencies present are within the observable region, and therefore we obtain a diffraction limited image as shown in Fig. 7.7b. We are imaging our sample with fluorescence, where the sample intensity distribution is a multiplication of the illumination distribution with the sample fluorescence distribution. If we multiplicatively superimpose on the ideal image from Fig. 7.7a, a sinusoidal pattern similar to that presented in Fig. 7.8a with a known frequency, τ, and phase, ϕ, we obtain $D(\boldsymbol{r})I(\boldsymbol{r})$ which is shown in Fig. 7.9a, where $D(\boldsymbol{r})$ is the underlying sample structure and $I(\boldsymbol{r})$ is the illumination pattern. From the convolution theorem we know that the Fourier transform of this image is a convolution of the image Fourier transform and the sinusoidal pattern Fourier transform, which results in an image with three identical copies of the same Fourier spectrum albeit it centred on different frequencies, as shown in Fig. 7.9b. We therefore obtain a resultant Fourier spectrum, $\widetilde{\mathbf{F}}'$ of the form:

$$\widetilde{\mathbf{F}}'(\boldsymbol{k}) = e^{-i\phi}\widetilde{\boldsymbol{D}}(\boldsymbol{k}+\tau) + \widetilde{\boldsymbol{D}}(\boldsymbol{k}) + e^{i\phi}\widetilde{\boldsymbol{D}}(\boldsymbol{k}-\tau), \tag{7.16}$$

where $\widetilde{\boldsymbol{D}}$ is the Fourier transform of the sample structure. However, the observable region still limits the spatial frequencies which can be captured by the imaging system, as shown in Fig. 7.9c. This results in a Fourier spectrum described by

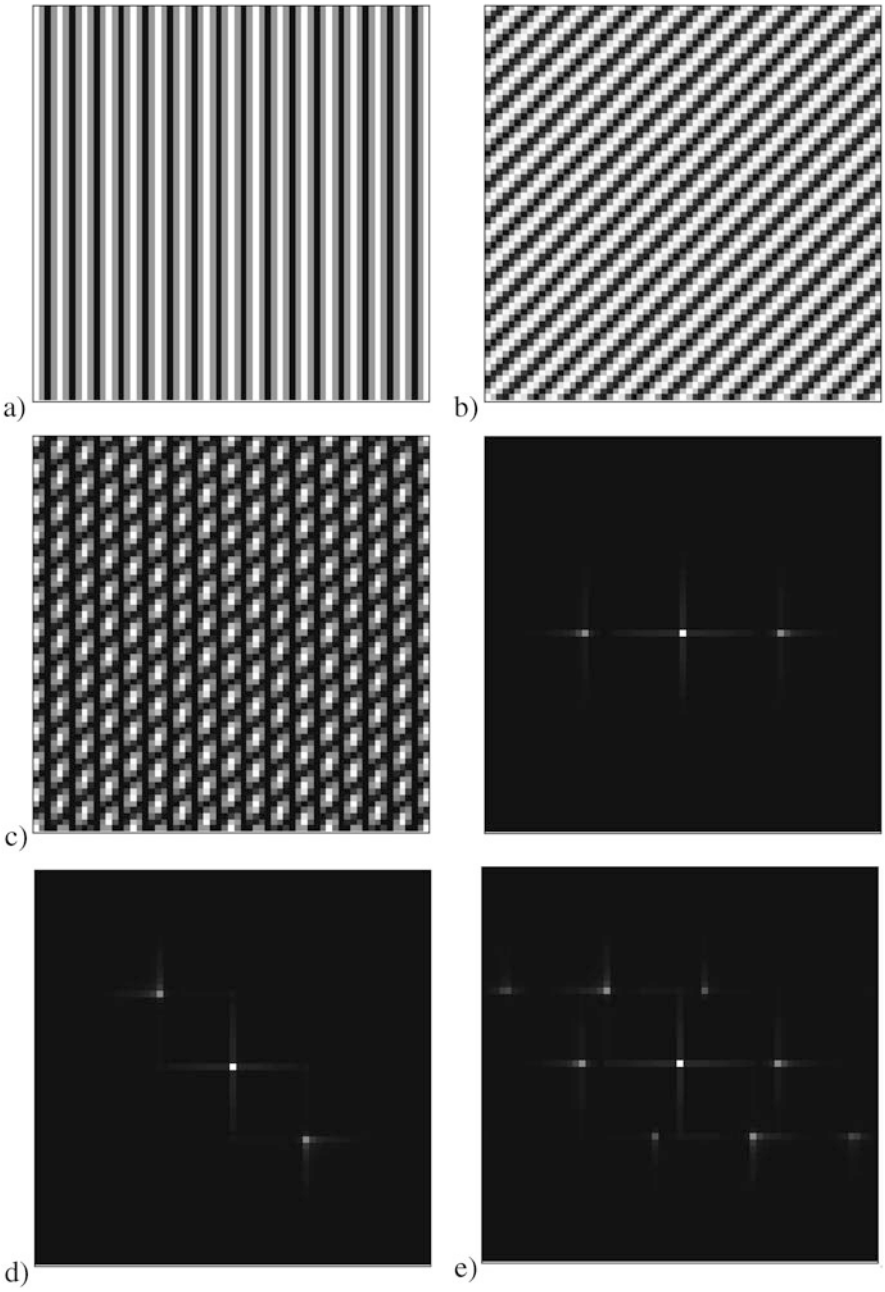

Fig. 7.8 Visualisation of moiré fringes. (**a**–**b**) Shows two spatially structured images. (**c**) Shows the resultant moiré fringes arising from the interference of (**a**, **b**). (**d**–**f**) Shows the respective Fourier transforms of (**a**–**c**). Note that (**f**) is a convolution of (**d**) and (**e**)

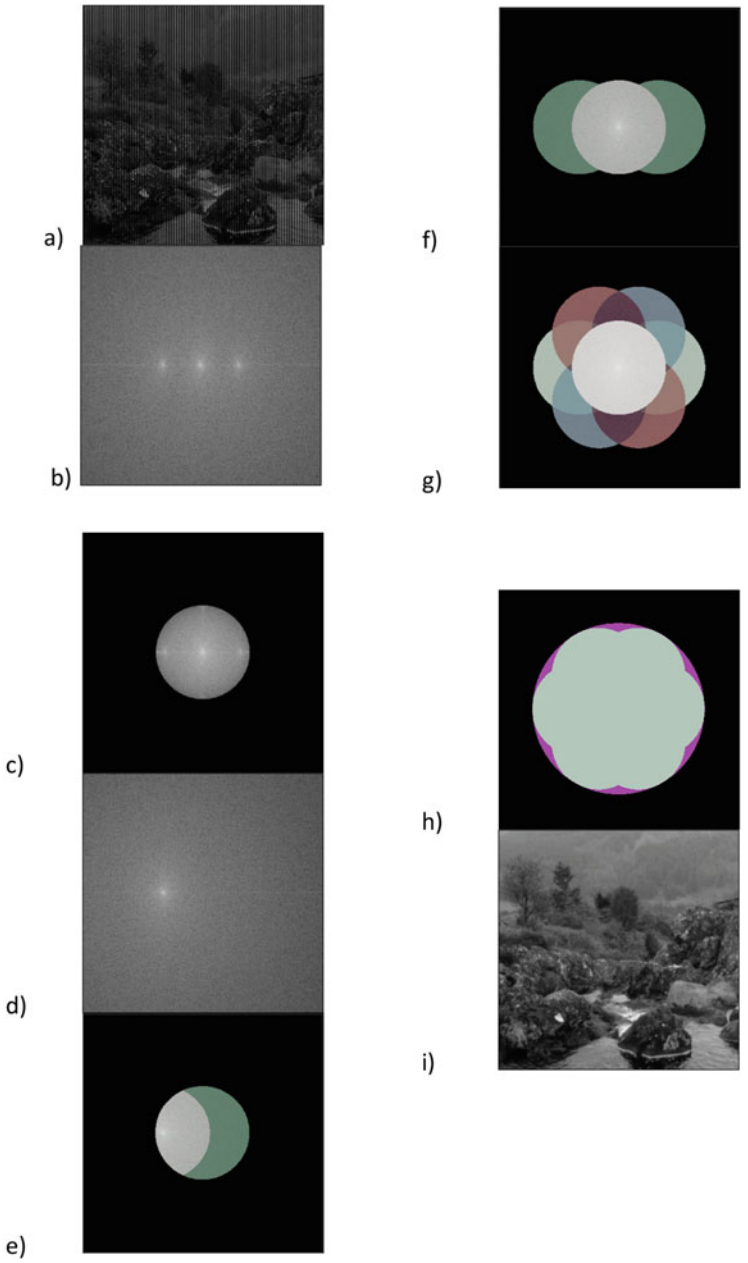

Fig. 7.9 Workflow of simulated 2D-SIM resolution enhancement. (**a**) The ideal image from Fig. 7.7a with a sinusoids pattern multiplicatively imposed on it. (**b**) Complete Fourier transform of (**a**). (**c**) The frequency components of the Fourier transform within the observable region. (**d**) Contribution to the Fourier transform in (**b**) arising from the convolution of the underlying image structure, $\widetilde{\boldsymbol{D}}(\boldsymbol{k})$, and the Fourier illumination component centred on $[k_x = \tau, k_y = 0]$. (**e**) The frequency components of the convolution within the observable region. Spatial frequencies which are normally outside of the observable region, the super-resolution spatial frequencies, are shaded in

$$\widetilde{\mathbf{F}}'(\boldsymbol{k}) = \widetilde{\boldsymbol{O}}_{-1}(\boldsymbol{k})e^{-i\phi}\widetilde{\boldsymbol{D}}(\boldsymbol{k}+\tau) + \widetilde{\boldsymbol{O}}_{0}(\boldsymbol{k})\widetilde{\boldsymbol{D}}(\boldsymbol{k}) + \widetilde{\boldsymbol{O}}_{1}(\boldsymbol{k})e^{i\phi}\widetilde{\boldsymbol{D}}(\boldsymbol{k}-\tau), \tag{7.17}$$

where $\widetilde{\boldsymbol{O}}_m$ is the optical transfer function of the imaging system, the Fourier transform of the point spread function (PSF). Let us now consider the contribution from the only one of the components, $e^{-i\phi}\widetilde{\boldsymbol{D}}(\boldsymbol{k}+\tau)$, which is shown in Fig. 7.9d. When we consider the region of this component which is within the observable region, Fig. 7.9e, clearly some spatial frequencies are ones which are already captured in conventional, widefield imaging. However, due to the lateral shift of the Fourier spectrum there are many spatial frequencies which were not previously within the observable region which are now captured by the imaging system, shown shaded in green in Fig. 7.9e. These spatial frequencies, beyond the resolution limit of the imaging system, contain the super-resolution information.

By isolating them and laterally shifting them to the correct position relative to the central contribution, the observable region is laterally extended. This enhanced resolution information is on and near to the line subtended by the Fourier components of the structured illumination pattern, shown shaded in green in Fig. 7.9f. To cover the whole range of angles the illumination pattern is rotated and the observable region can be extended in multiple directions, shown shaded in Fig. 7.9g. Rotating the illumination pattern such that the various Fourier components are spaced $\frac{\pi}{3}$ from one another gives a near isotropic improvement in lateral resolution. The comparison between a true isotropic doubling of the resolution limit and the extended observable region created by 2D-SIM is shown in Fig. 7.9h, with the difference in observable region extent shaded in magenta. Figure 7.9i shows a 2D-SIM reconstruction of the diffraction limited image shown in Fig. 7.7b. Clearly Fig. 7.9i recovers many of the fine details which are unobservable in Fig. 7.7b and appears far closer to the ideal image shown in Fig. 7.7a.

Readers may be wondering how the contributions from the 3 Fourier components can be separated. This problem is actually relatively simple to understand. If we consider any single Fourier spatial frequency within the observable region $\boldsymbol{k}_i$, then according to Eq. (7.17), the amplitude at the frequency $\boldsymbol{k}_i$ is a sum of three contributions. In other words, it is a linear equation with 3 unknowns. As is well documented, if one has N linear equations with N unknowns, then one can solve this system of linear equations and obtain the value for each of the unknowns. Since each of the contributions has not only a phase component, $e^{im\phi}$, but this phase component is different for each contribution, by varying the phase of the structured illumination

Fig. 7.9 (continued) green. (**f**) The extended observable region when the super-resolution spatial frequencies are laterally shifted to their correct locations. (**g**) The extended observable region when sinusoidal patterns of multiple direction are imposed on the ideal image, imaged through the observable region and the super-resolution frequencies are relocated to their correct position. (**h**) Comparison between a true isotropic doubling of the resolution limit and the extended observable region shown in (**g**). The regions where the extended observable region does not match the doubled resolution limit are shown in purple. (**i**) A 2D-SIM reconstruction of the diffraction limited image shown in Fig. 7.7b, imaged through the extended observable region

pattern for different images it is possible to solve this system of linear equations for each spatial frequency and thereby determine the contribution of each Fourier component. Since the contributions are laterally translated by a known amount, $\pm\tau$, it is possible to translate them to their correct position, leading to an extended observable region.

From this explanation it should be intuitive why, for each angle we wish to extend the observable region in 2D-SIM, we require 3 images. However, for 3D-SIM we have 7 Fourier components and so it might seem like there has been an error when we say we only need 5 images per angle. Here, we utilise the property of the Fourier transform of real functions described in Eq. (7.3) and can conclude that the $+\boldsymbol{k}_z$ contributions are identical to the $-\boldsymbol{k}_z$ contributions. We therefore only have 5 unknowns rather than 7, and therefore only need 5 images to solve this system of linear equations.

7.4.1 Reconstructing a SIM Image

The general process of reconstructing a SIM image from multiple raw widefield images takes several steps, described below but there are a few general points to be made first. The exact mathematical description of the reconstruction will not be covered here, but it is readily available in the literature and there a number of open-source implementations [10–14].

The SIM reconstruction relies upon information shifted in frequency by the stripes (or other patterns) in the illumination, then propagated into the recorded image. The most critical factor in gaining good SIM reconstructions is to have high contrast in the original illuminating stripes and in the produced fluorescent images. This is influenced by both the optical setup and the sample. Imaging at depth with SIM is challenging as sample-induced aberration, usually spherical aberration, reduces and finally eliminates stripe contrast as you image deeper into the sample. Two common techniques to reduce, or eliminate spherical aberration, are to vary the refractive index of the immersion oil or to use an objective with a correction collar. Either of these techniques can dramatically improve stripe contrast and hence reconstruction quality. However, it should be noted that mismatch between the immersion and sample refractive indices often mean this correction only applies over a limited depth range. Dispersion, variation of the refractive index with wavelength, in samples can also mean the required correction differs at different wavelengths. Additionally imaging samples with high out of focus background is much more challenging than samples with very low background.

The steps in the SIM reconstruction process are:

1. The images are converted into reciprocal space using a Fourier transform.
2. The multiple images at a single orientation are used as a set of linear equations to extract the components due to the different Fourier elements in the excitation light.

3. The components are then divided by the optical transfer function to normalise the amplitude of information at different spatial frequencies as the microscope detected the different spatial frequencies with a varying sensitivity.
4. The normalised components are then shifted to their correct positions and combined to generate an extended resolution image in reciprocal space. The exact shift used is usually optimised on an image by image basis by finding the highest correlation between the expected overlapping regions in the images from the multiple components. This allows correction for errors introduced by small changes in the optical setup due to issues such as temperature variations.
5. Steps 2-4 are repeated for each orientation. As previously mentioned, often there are three orientations as this minimises the number of images required while producing an almost isotropic result.
6. All orientations are combined to produce a single reciprocal space image.
7. The combined reciprocal image then has a frequency based filter applied, usually a Weiner filter is used.
8. Finally the reciprocal space image is inverse Fourier transformed to generate the final reconstruction.

The spatial filter has two effects, it produces an image which is comparable to other microscope images, it effectively reverses the division by the OTF which was applied earlier, but with a synthetic OTF which has twice the frequency limit, as the image has double the resolution. Secondly, the filter strongly suppresses high frequency noise that is amplified in the initial processing as the original noisy data is divided by the OTF. The OTF has low amplitude near the edge of the observable region meaning noise here contributes more to the final image.

In 2D SIM the above processes are performed in 2D, whereas in 3D SIM the processing uses 3D images throughout. In 3D SIM this means that the processing can account for the out of focus contributions and properly double the resolution in Z as well as in the XY directions. This does depend on the original 3D image stack including in focus information for all significant contributions. Meaning thicker samples may require large stacks to ensure that all significant fluorescence has both in focus and out of focus information included.

The initial stages of the processing are entirely linear and well constrained. The final filtering step does introduce a non-linear element but is extremely important to suppress high frequency noise components. The effective cut-off frequency of the final filtering can be tuned to play off resolution against noise, with stronger filtering eliminating more high frequency noise at the cost of lower resolution in the final reconstruction.

The output reconstruction quality depends on many factors and data quality should be carefully checked to ensure that experimental conclusions are adequately supported by the imaging data. There are resources for sample preparation, imaging and data quality assessment available [15–17].

7.5 Strengths and Limitations of SIM

Each super-resolution technique has benefits and drawbacks which need to be considered when employing them including; the resolution required (both laterally and axially), the acquisition speed, photodamage incurred, depth of imaging, and multi-colour capability [16, 18].

Although it has a modest resolution improvement—only $\times 2$ compared to conventional microscopy much less than stimulated emission depletion (STED) microscopy and single-molecule localisation microscopy (SMLM) which offer many times that, SIM a widely applied super-resolution technique for biology for a number of reasons [16, 19]. The number of images required in order to reconstruct a super-resolution SIM image is typically 9 for 2D SIM and 15 for 3D SIM, determined by the number of lateral Fourier components in the illumination pattern—3 for 2D SIM and 5 for 3D SIM—and the number of stripe angles used to rotate the illumination pattern through, typically 3. This relatively low number of images per reconstruction has a number of benefits. Firstly, the temporal resolution is considerably higher for SIM than for point-scanning super-resolution techniques, such as STED, or SMLM techniques which require a great deal more images per reconstruction [16, 20]. Secondly, each of these images is fundamentally still a widefield-style image and therefore has a lower light-dosage per image than STED or SMLM techniques. The lower light-dosage per image combined with fewer total images required per reconstruction contributes to a low photodamage impact. Finally, SIM is easily expandable to multi-colour imaging using a wide range of standard fluorophores not requiring specific dyes, or imaging buffers [21, 22]. Overall, SIM is a super-resolution technique well suited to biological imaging, particularly imaging dynamic biological processes.

Take Home Message

- Structured illumination microscopy is one of the most widely used super-resolution techniques in modern microscopy.
- The low-light dosage, low number of images required to achieve super-resolution and multi-colour extensibility make it well suited to biological imaging, particularly in live samples.
- The quality of your final super-resolution images is critically dependent on having a high-contrast structured illumination pattern in the images used for the reconstruction.
- The resolution improvement relative to other super-resolution techniques is modest.

References

1. Körner TW. Fourier transforms. In: Fourier analysis. Cambridge: Cambridge University Press; 1988. p. 219–20.

2. Rayleigh L. XII. On the manufacture and theory of diffraction gratings. Lond Edinb Dublin Philos Mag J Sci. 1874;47(310):81–93.
3. Rayleigh L. Investigations in optics, with special reference to the spectroscope. Mon Notices R Astron Soc. 1880;40:254.
4. Pawley J. Handbook of biological confocal microscopy, vol. 236. New York: Springer; 2006.
5. Roy Frieden B. Optical transfer of the three-dimensional object. JOSA. 1967;57(1):56–66.
6. Abbe E. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. Archiv für mikroskopische Anatomie. 1873;9(1):413–8.
7. Behan G, et al. Three-dimensional imaging by optical sectioning in the aberration-corrected scanning transmission electron microscope. Philos Trans R Soc A. 2009;367(1903):3825–44.
8. Arnison MR, Sheppard CJR. A 3D vectorial optical transfer function suitable for arbitrary pupil functions. Opt Commun. 2002;211(1–6):53–63.
9. McGillem CD, Cooper GR. Continuous and discrete signal and system analysis. Oxford: Oxford University Press; 1991.
10. Gustafsson MGL. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. J Microsc. 2000;198(2):82–7.
11. Gustafsson MGL, et al. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. Biophys J. 2008;94(12):4957–70.
12. Müller M, et al. Open-source image reconstruction of super-resolution structured illumination microscopy data in ImageJ. Nat Commun. 2016;7(1):1–6.
13. Lal A, Shan C, Xi P. Structured illumination microscopy image reconstruction algorithm. IEEE J Sel Top Quantum Electron. 2016;22(4):50–63.
14. O'Holleran K, Shaw M. Optimized approaches for optical sectioning and resolution enhancement in 2D structured illumination microscopy. Biomed Opt Express. 2014;5:2580–90.
15. Ball G, et al. SIMcheck: a toolbox for successful super-resolution structured illumination microscopy. Sci Rep. 2015;5:15915.
16. Schermelleh L, et al. Super-resolution microscopy demystified. Nat Cell Biol. 2019;21(1):72.
17. Spilger R, et al. Deep probabilistic tracking of particles in fluorescence microscopy images. Med Image Anal. 2021;72:102128.
18. Hell SW, et al. The 2015 super-resolution microscopy roadmap. J Phys D Appl Phys. 2015;48 (44):443001.
19. Schermelleh L, Heintzmann R, Leonhardt H. A guide to super-resolution fluorescence microscopy. J Cell Biol. 2010;190(2):165–75.
20. Leung BO, Chou KC. Review of super-resolution fluorescence microscopy for biology. Appl Spectros. 2011;65(9):967–80.
21. Wu Y, Shroff H. Faster, sharper, and deeper: structured illumination microscopy for biological imaging. Nat Methods. 2018;15(12):1011–9.
22. Allen JR, Ross ST, Davidson MW. Structured illumination microscopy for superresolution. ChemPhysChem. 2014;15(4):566–76.

Further Reading

Prior Knowledge

Hecht E. Optics. 4th ed. Cambridge: Addison Wesley Longman; 1998.
Illing L. Fourier analysis. J Pure Appl Math Ser. 2008;7:1–21.
Morin D. Fourier analysis. Cambridge: Harvard University; 2009.

Mathematical Derivations of 2D and 3D SIM

Gustafsson MGL. Extended resolution fluorescence microscopy. Curr Opin Struct Biol. 1999;9(5):627–8.

Gustafsson MGL. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. J Microsc. 2000;198(2):82–7.

Gustafsson MGL, et al. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. Biophys J. 2008;94(12):4957–70.

Heintzmann R, Cremer CG. Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating. Opt Biopsies Microsc Tech III. 1999;3568:185–96.