3D Light Lattice Structured Illumination Microscopy

Lucas Rietjens*

June 20, 2026

Delft University of Technology
Double Bachelor Thesis
Applied Physics and Applied Mathematics

Supervisors:
Prof. Dr. S. Stallinga (TU Delft, Faculty of Applied Sciences)
Dr. P.M. Visser (TU Delft, Faculty of EEMCS)

Layman Summary

Traditional microscopes are limited by the laws of physics regarding how far they can zoom in on details before the images become blurred. Structured illumination microscopy (SIM) is a modern imaging technique that uses mathematical tools to work around the limitations. By shining highly detailed patterns of light onto the object of interest, such as a living cell, SIM can reveal tiny hidden details that would otherwise be blurred.

To utilise these mathematical tools and produce a final clear image, computers must run a series of computational steps. Scientists working with SIM often prioritise the speed of image acquisition, and therefore within these calculations some complexity is neglected. This helps to save time, however it comes with the drawback that information about the image is lost.

The skipped step concerns the interaction between the light beams that are shone on the object. Scientists work under the assumption that these beams do not interfere with each other, however in reality they do. This thesis investigates the degree to which taking this shortcut ruins the final image quality. This is done by testing different shapes of light grids. The research concluded that for the commonly used setups, the assumption step works well enough that prioritising the imaging speed is more beneficial than the extra image quality. However, for very crowded light patterns, the conclusion is the opposite, as then the light interferes more significantly and it becomes more penalised to work with the shortcut, sometimes being more than 50% worse than proper computation.

Abstract

Structured illumination microscopy (SIM) is a super resolution technique that bypasses the traditional diffraction limit. This is achieved by shining a spatially periodic illumination pattern onto a sample, which reveals finer details because high spatial frequencies can be downshifted within the optical transfer function (OTF) support of the imaging system. During the image reconstruction process, it is the convention to neglect the so-called noise matrix. The noise matrix describes the amount of cross-talk between Fourier orders, however, the simplification ignores the cross-talk by assuming that this matrix can be approximated by $M = I$. This assumption is made to reduce the computational load and processing time, but it comes at the potential cost of image quality.

This thesis studies the validity of the suboptimal filter assumption for a range of light lattice configurations, including conventional SIM, plus shaped light lattices and unit cell grid layouts. By comparing analytical bounds, such as the Kantorovich inequality and the Bounding Theorem, against exact spectral signal-to-noise ratio (SSNR) numerical calculations, this research has concluded that analytical bounds are often too loose to properly represent the physical system. Consequently, numerical analysis is done. The numerical results are that for sparse lattice configurations, such as 2D conventional SIM and plus shaped light lattices, the simplification is highly valid since the noise matrix is strongly diagonally dominant. This means there is minimal penalty and the SSNR ratio remains above 0.8 across all spatial frequencies. However, for dense lattices, such as high pitch unit cell grids, significant cross-talk occurs between orders. This invalidates the simplification with image quality penalties of upwards of 50% to 60% near cutoff. Furthermore, as the lattice pitch approaches infinity, the system converges to the image scanning microscopy (ISM) limit, where using the exact noise matrix becomes essential to obtaining image clarity.

Contents

1 Introduction

Producing images of fine, highly-detailed structures is a challenging task within the field of imaging physics. Using traditional light microscopy techniques, the production of images requires light photons to be detected by a lens. This restricts the resolution, since traditional imaging techniques face the limitation called the Abbe diffraction limit, $d = \frac{\lambda}{2NA}$ [1]. This imposes a resolution limit that claims that details can only be distinguished if they are separated by a minimum distance that depends on the numerical aperture (NA) of the lens and wavelength of the light used. This is illustrated in Figure 1. The numerical aperture is defined as the product of half of the cone angle of light that passes into the lens objective, $\sin\theta$, and the refractive index, n, of the medium between the specimen and lens, $\mathrm{NA} = n\sin\theta$ [2]. The NA describes the light gathering ability of an objective lens, shown visually in Figure 2. The frequently quoted resolution limit for visible light with a high NA lens is 200 nm [3]. Modern physicists have derived numerous super resolution techniques [4] that are able to bypass the diffraction limit. One such example is structured illumination microscopy (SIM) [5].

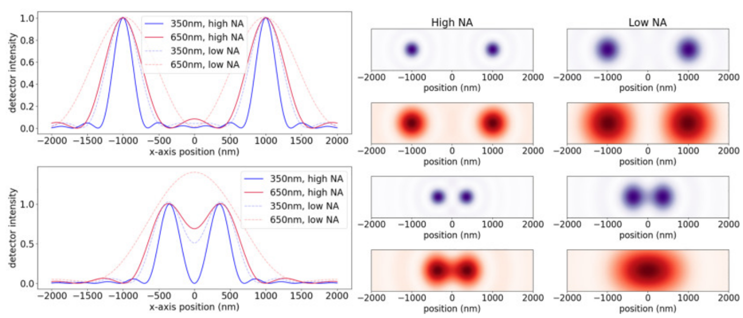

Figure 1: This image visually shows the Abbe diffraction limit for two different NA lenses and wavelengths of light. Two point sources give off light that interferes. The measured intensity pattern blurs together as the two sources are brought closer together, making the peaks harder to differentiate. The Abbe diffraction limit defines the minimum distance between the two sources such that they can be individually resolved. This image displays the significance of the NA and the wavelengths on this minimum distance. Image sourced from [6].

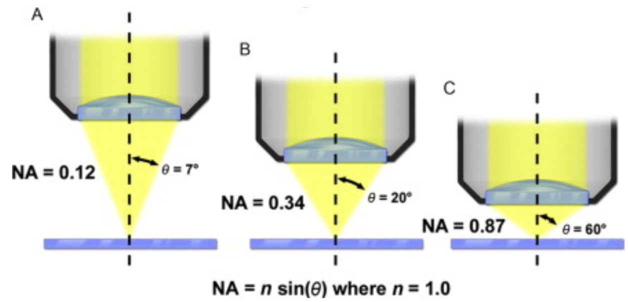

Figure 2: Illustration of the relationship between the collection half angle, numerical aperture, and working distance of a microscope objective. (A) A long working distance objective, with a correspondingly small collection half angle. (B) A shorter working distance, higher numerical aperture, objective. (C) A short working distance, high numerical aperture objective. Image and caption sourced from [2].

Structured illumination microscopy is a relatively new imaging technique that was first theorised by Lukosz and Marchand in 1963 [7], but it was only in the late 1990's that practical uses for SIM were demonstrated [8]. SIM is a super resolution technique that is able to bypass the diffraction limit of regular microscopy. This is done using the Moiré effect [9], where by taking two highly detailed patterns and overlapping them, they create a new interference pattern composed of Moiré fringes. SIM applies this by shining a spatially periodic, non-uniform illumination pattern on a sample. This illumination pattern overlapping with the sample pattern will produce Moiré fringes that can be imaged to observe fine details since the high spatial frequencies have been downshifted within the band limit. Figure 3 displays how SIM can be used to achieve images of a highly detailed object without blur.

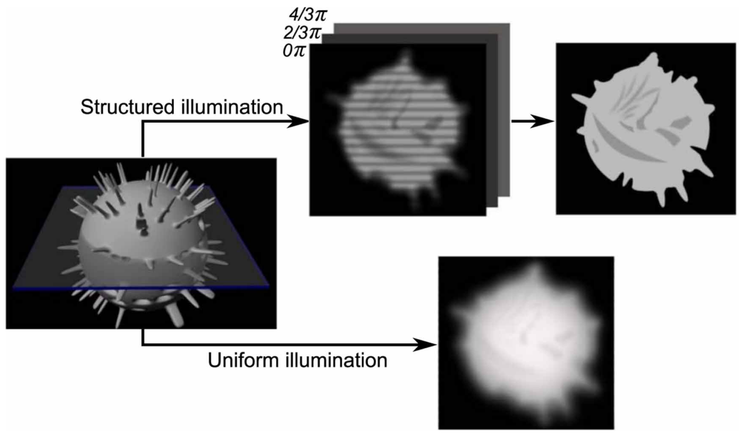

Figure 3: The principle of optical sectioning SIM. Optical sectioning is achieved by translating the structured illumination, out-of-focus blur is effectively rejected because it is not modulated by the SIM pattern. Image sourced from [10]. Caption sourced from [11].

A number of super resolution techniques have been developed, each of which has applications in different imaging tasks [12]. Through the use of SIM, scientists can study tiny samples such as biological cells in more detail, specifically for live-cell imaging [13]. Unlike materials, cells are living organisms. Alternative imaging techniques such as scanning electron microscopy or using x-rays use lower wavelength radiation and can thus have far greater resolutions, however, these are not suitable for use on live cells ([14], [15]). This is because these tools pose the risk of damaging a live sample through radiation damage with x-rays or through the vacuum environment required for scanning electron microscopy [16]. Therefore, imaging live cells restricts the electromagnetic radiation to visible light to ensure that the sample stays alive. Moreover, this issue of live-cell imaging requires that images can be taken and processed very quickly. An example of images taken of live cells using SIM is shown in Figure 4. Historically, SIM had used illumination patterns that were stripes of light, fringes, to produce a periodic sinusoidal pattern, where images are produced by translating and rotating these stripes [17]. This project will look at light lattices rather than stripes of light, where to produce images this light lattice only needs to be translated. This can save a significant amount of time during imaging and hence improve live-cell imaging or other imaging fields.

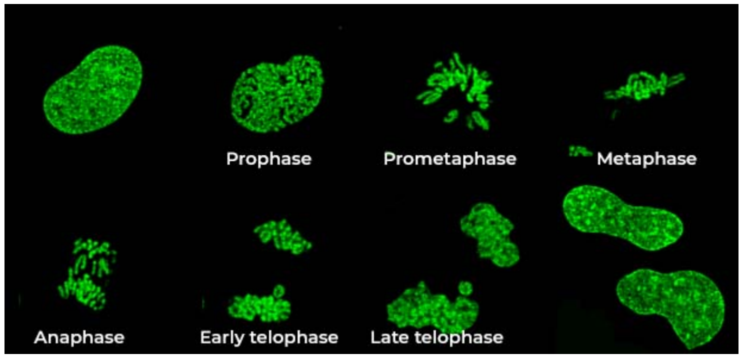

Figure 4: This figure shows the different mitotic phases of a cellular division, imaged using SIM. These images were acquired with CrestOptics X-Light-V3 CF spinning disk system coupled with DeepSIM SR add-on and equipped with CFI Plan Apo Lambda 60X oil objective. Image sourced from [18].

It is standard practice to make assumptions or simplifications when the consensus is that the error that results from this decision is negligible. SIM is no exception to this. For images to be produced using SIM, a reconstruction algorithm is used to piece together different elements into a final super resolved image [19]. During this process, the convention is to make a simplification step in the algebra, assuming that a diagonally dominated matrix M can be assumed to be a scaled identity matrix. This assumption results in a less optimal final image, at the benefit of reduced computation. This thesis is going to assess the possible improvements that can be made to the final image quality by rigorously including the proper form of M during reconstruction. The potential improvement will be assessed by comparing the spectral signal-to-noise ratio (SSNR) achieved by both approaches.

The matrix M is often referred to as the noise matrix, and the off-diagonals are the terms due to interference between light sources in the lattice, called cross-talk. The convention to ignore the cross-talk between orders has variable impact depending on the type of lattice used. This project will also aim to assess if the improvement is more noticeable for different types of lattices to eventually conclude whether this additional computation step should not be neglected during reconstruction.

2 Theory

2.1 The Concept of SIM

In standard imaging, there are limitations to the maximum obtainable resolution. When considered in the Fourier domain (often interchanged with frequency domain and k-space throughout this paper), a microscope lens acts as a low-pass filter that can only capture spatial frequencies of light up to a certain cutoff, which is determined by its optical transfer function (OTF). The OTF represents the ability of the optical system to transfer different spatial frequencies. The OTF typically decays smoothly until reaching zero at the cutoff frequency, beyond which no light is transferred. During image acquisition, all higher frequency information above the cutoff frequency is lost, $f_{\text{cutoff}} = \frac{2NA}{\lambda}$. This effect is observed in Figure 5, which shows the low-pass filter effect of standard imaging in both real space and in the Fourier domain. The top half of this figure depicts the real space. In this domain, the image is formed through a convolution between the real image and the point spread function (PSF), in this example a sinc function. The image loses clarity compared to the true image representation. The bottom half shows the same process in the Fourier domain. In this domain, the full Fourier representation of the image is multiplied by the OTF. Here, the OTF shape represents a low-pass filter, which filters out the higher spatial frequencies.

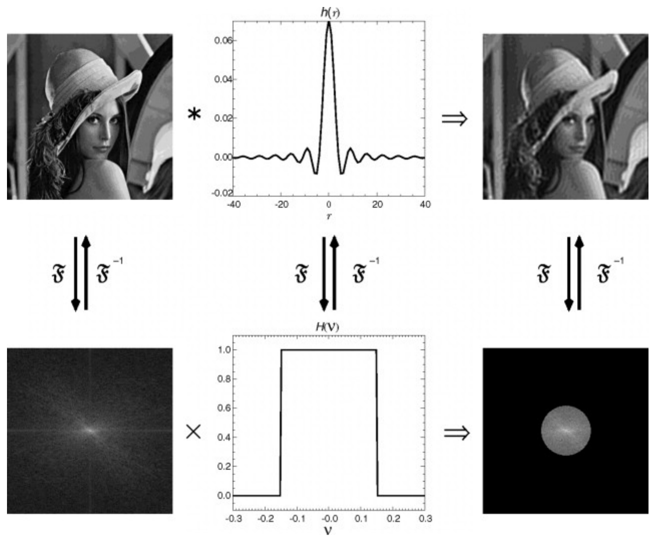

Figure 5: This figure shows the digital imaging process in both the real domain (top half) and in the Fourier domain (bottom half). In the real domain the image is formed through a convolution with a sinc PSF function, resulting in lost details. These lost details can be seen in the Fourier representation, where a low-pass filter representing the imaging system, the OTF, removes higher spatial frequency details. Image sourced from [20].

Structured illumination microscopy aims to capture these higher spatial frequencies by downshifting them within the cutoff, artificially increasing the width of the OTF. This provides enhanced lateral resolution since it allows for the reconstruction of higher frequencies. The downshifting process is done by utilising the Moiré effect. This effect is observed when two periodic patterns are overlapped (mixed) with each other, resulting in the emergence of a new pattern. This is visually shown in Figure 6. Mathematically, this downshifting occurs because the periodic patterns can be represented as sinusoidal waves. When these patterns overlap, their physical multiplication can be rewritten using the addition formulae for trigonometric functions. This transforms the product of two cosines into a sum of two new cosines at new frequency bands, one of which is a lower, measurable spatial frequency, as shown in Equation 1.

$$\cos(2\pi\vec{q}_1 \cdot \vec{r})\cos(2\pi\vec{q}_2 \cdot \vec{r}) = \frac{1}{2}\cos(2\pi(\vec{q}_1 + \vec{q}_2) \cdot \vec{r}) + \frac{1}{2}\cos(2\pi(\vec{q}_1 - \vec{q}_2) \cdot \vec{r}) \quad (1)$$

In this expression the spatial frequency is denoted by $\vec{q}$. This is called frequency modulation [21]. This modulation has therefore downshifted some of the information to a lower spatial frequency, within the support region of the system. By knowing the characteristics of the projected illumination patterns, it is possible to unmix the frequencies and recover information about the sample. ([22], [23])

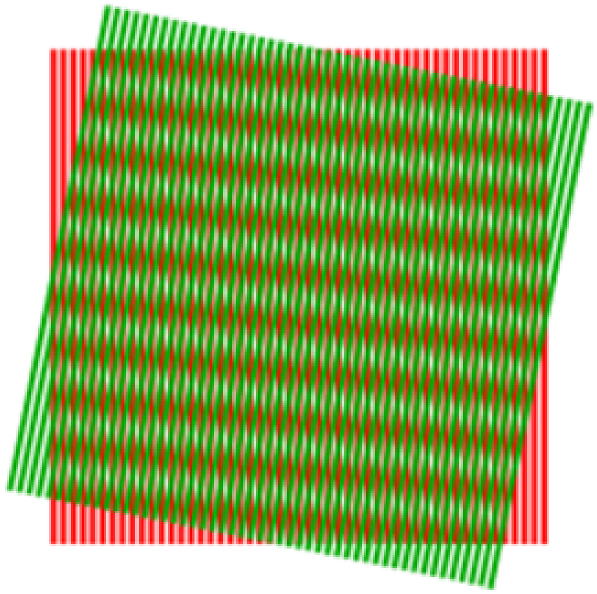

Figure 6: An illustration of two overlapping patterns, which produce a new pattern. This is the Moiré effect. Figure sourced from [21].

The resulting image after this downshifting is a mixture of the normal low-frequency image and the shifted high-frequency information. To untangle these, multiple images are taken by translating the illumination pattern. This shifts the phase of the fringes. Using a sequence of images at different phases then allows for the construction of a system of linear equations to computationally separate the frequency bands. Once separated, these frequency bands are computationally shifted back to their correct positions in Fourier space and summed up together to reconstruct a single, higher resolution image. These steps are visually shown in Figures 7 and 8. The method of SIM allows for an improvement of up to 2x as compared to standard microscopy [21]. Figure 9 shows the difference between an image taken with SIM and with wide-field. In wide-field the specimen is illuminated all at once, which is limited by the standard diffraction limit, resulting in a more blurred image. This image illustrates the improvement in detail resolved using SIM.

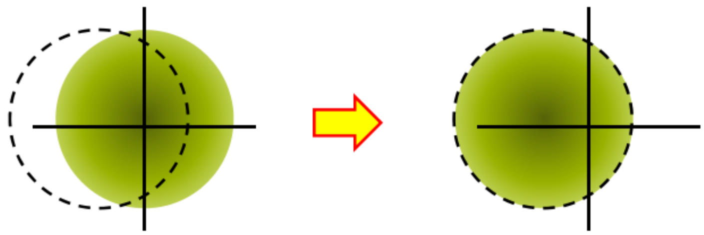

Figure 7: An illustration of how downshifted information (filled green circle) is shifted from the downshifted location in the origin where it can be measured, back to its original position in Fourier space (dotted circle). The circular shape represents the support region of the microscope OTF as seen prior in Figure 5. Figure sourced from [21].

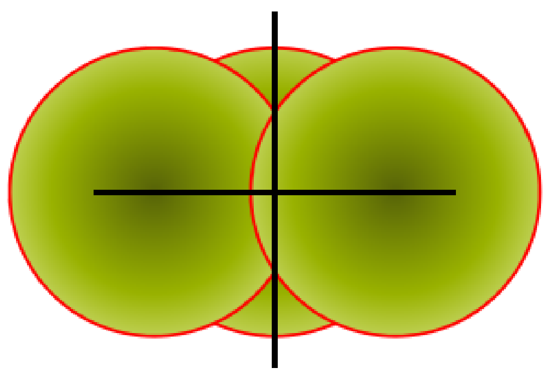

Figure 8: An illustration of how information from multiple images of shifted information is summed together to form a single high resolution image that includes the larger spatial frequencies. This shows how the effective OTF is extended using SIM. Figure sourced from [21].

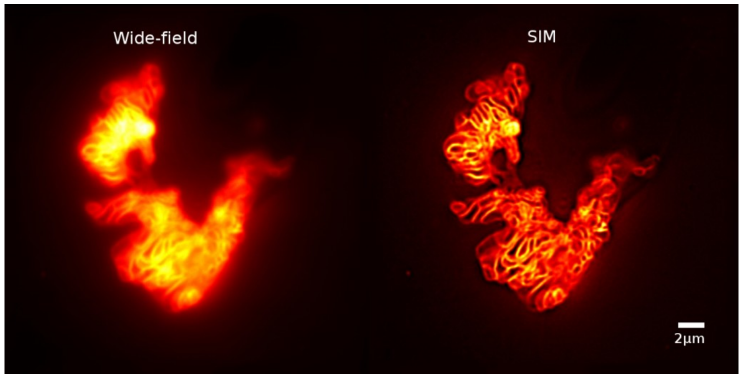

Figure 9: GFP-stained mitochondria in COS-7 cells imaged with conventional wide-field and structured illumination microscopy. (Sample from Laura Osellame and Mike Ryan, La Trobe University.) Image and caption sourced from [24].

2.2 Illumination Patterns

One of the most vital elements of lattice SIM is the illumination pattern that is used to illuminate the sample. The shape of the light lattice influences the illumination pattern created. Consequently the light lattice must be chosen such that the resulting illumination pattern is periodic. In this section, various possible light lattice structures and their resulting illumination pattern are discussed. The derivations of the amplitudes for the Fourier orders of different light lattice types will be explained. The results can be found directly in the Appendix in Tables 1 to 5. Note that the amplitude ratio has been taken to be $r = 1$, meaning each beam has the same amplitude. Representations of commonly used SIM configurations can be seen in Figure 10.

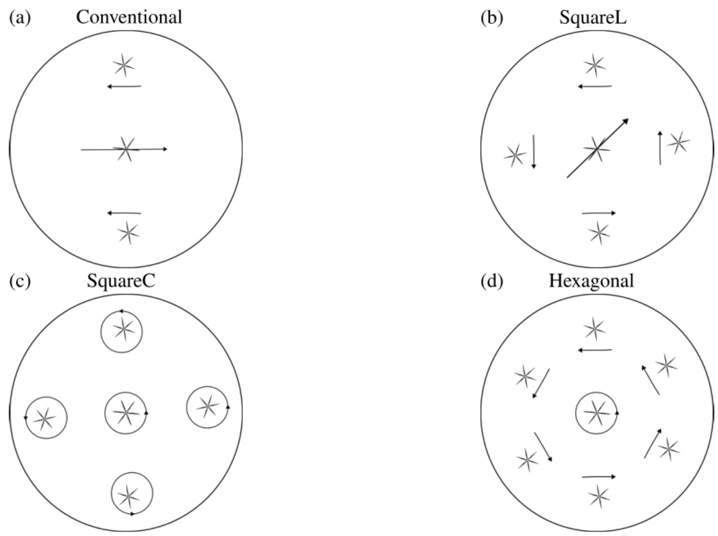

Figure 10: Back focal plane illumination configurations for different 3D SIM configurations. a) conventional 3D SIM ("Conventional"). b) Square lattice 3D-SIM with linearly polarised plane waves ("SquareL"). c) Square lattice 3D SIM with circularly polarised plane waves ("SquareC"). d) Hexagonal lattice SIM with linearly polarised plane waves ("Hexagonal"). Image and caption sourced from [25].

The illumination pattern is denoted by $h(\vec{r})$ and is related to the electric field by Equation 2 [25].

$$h(\vec{r}) = |\vec{E}(\vec{r})|^2 \tag{2}$$

The electromagnetic field is created by superimposing multiple plane waves generated by a laser. Note, that when working with lattices there are multiple contributing plane waves, and hence the resulting total field will contain a sum of many terms. This can make this calculation tedious and potentially time consuming. Consequently, it is advantageous to determine this illumination pattern directly within the Fourier domain.

In the Fourier domain the sources are represented as delta peaks at the source position in k-space. The illumination pattern can then be derived by computing the autocorrelation of these delta peaks. The formula for the autocorrelation is shown in Equation 3, where $\hat{E}(\vec{k})$ is the Fourier transform of the electric field.

$$(\hat{E} \star \hat{E})(\vec{q}) = \int \hat{E}(\vec{k})\hat{E}^*(\vec{k} - \vec{q})d\vec{k} \tag{3}$$

2.2.1 1D Light Lattice Types

A 1D light lattice is a lattice that contains light sources that are aligned along the same axis in the back focal plane (BFP). The to be discussed light lattice types in this section are the conventionally used lattices since they are computationally easy, yet they are able to achieve super resolution. The most simple light lattice used in SIM is 2D conventional SIM. This light lattice is made up of two delta peaks in Fourier space. To maximise the interference between the two beams, they are both s-polarised (linearly polarised along the x-axis). Let the two beams be located at $\vec{k}_y$ and $-\vec{k}_y$ along the y-axis, then the resulting electric field is given in Equation 4.

$$\vec{E}(\vec{r}) = A\left(e^{i2\pi\vec{k}_y\cdot\vec{r}} + e^{-i2\pi\vec{k}_y\cdot\vec{r}}\right)\hat{x} \tag{4}$$

In the Fourier domain, the electric field is then given by Equation 5.

$$\hat{E}_x(\vec{k}) = A\delta(\vec{k} - \vec{k}_y) + A\delta(\vec{k} + \vec{k}_y) \tag{5}$$

Computing the autocorrelation, $\hat{h}(\vec{q}) = (\hat{E}_x * \hat{E}_x)(\vec{q})$, this gives three distinct Fourier orders at the positions $\vec{q} \in \{0, 2\vec{k}_y, -2\vec{k}_y\}$. The final illumination pattern is given by Equation 6.

$$\hat{h}(\vec{q}) = 2A^2\delta(\vec{q}) + A^2\delta(\vec{q} - 2\vec{k}_y) + A^2\delta(\vec{q} + 2\vec{k}_y) \tag{6}$$

This illumination pattern in Fourier space is shown in Figure 11 and in real space the periodic pattern is shown in Figure 12.

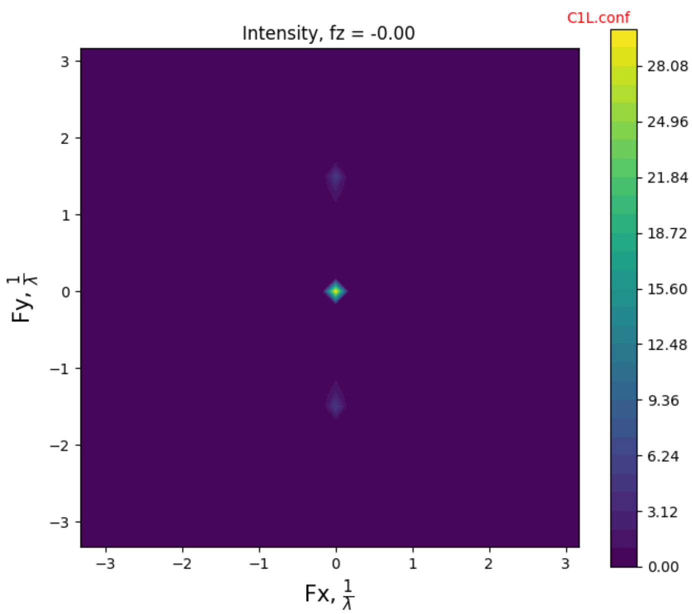

Figure 11: The resulting illumination pattern in 3D Fourier space of the 2D conventional SIM light lattice, viewed at an axial slice of $f_z = 0.00$. Figure generated with code from [26], where $\vec{k}_y = (0, 0.75)$. The colour bar on the right indicates the intensity of the illumination pattern at a given point, using $A = 3$.

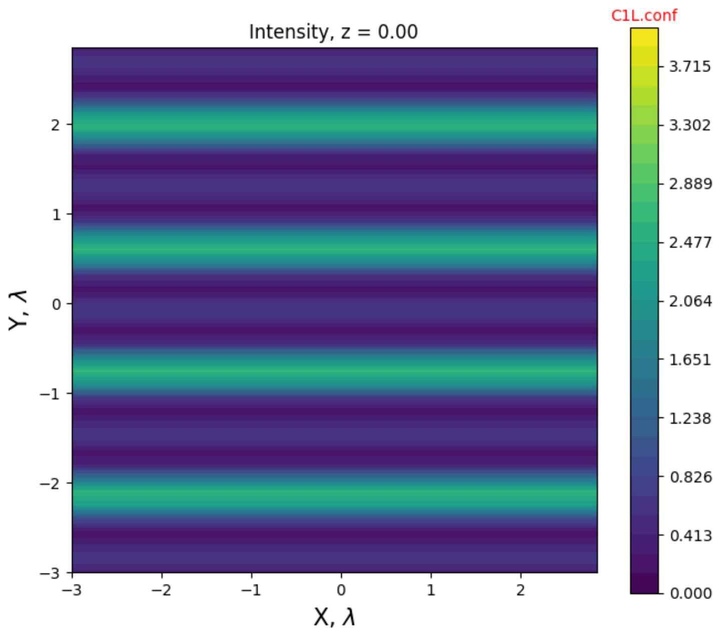

Figure 12: The resulting illumination pattern in 3D real space of the 2D conventional SIM light lattice, viewed at an axial slice of $f_z = 0.00$. Figure generated with code from [26], where $\vec{k}_y = (0, 0.75)$. The colour bar on the right indicates the intensity of the illumination pattern at a given point, using $A = 3$.

To achieve optical sectioning, a third beam is introduced resulting in the light lattice known as 3D conventional SIM. In this light lattice, three delta peaks in Fourier space lie on the same axis in the BFP, seen in Figure 10(a). These peaks are located at positions $\vec{q} \in \{\vec{0}, \vec{k}_y, -\vec{k}_y\}$, with $\vec{k}_y = (0, k_y)$. The electric field is then given by Equation 7.

$$\vec{E}(\vec{r}) = A(1 + e^{i2\pi\vec{k}_y\cdot\vec{r}} + e^{-i2\pi\vec{k}_y\cdot\vec{r}})\hat{x} \tag{7}$$

The Fourier transform of this gives the result in Equation 8.

$$\hat{E}_x(\vec{k}) = A\delta(\vec{k}) + A\delta(\vec{k} - \vec{k}_y) + A\delta(\vec{k} + \vec{k}_y) \tag{8}$$

Computing the autocorrelation then gives the illumination pattern for 3D conventional SIM shown in Equation 9.

$$\hat{h}(\vec{q}) = 3A^2\delta(\vec{q}) + 2A^2[\delta(\vec{q} - \vec{k}_x) + \delta(\vec{q} + \vec{k}_x)] + A^2[\delta(\vec{q} - 2\vec{k}_x) + \delta(\vec{q} + 2\vec{k}_x)] \tag{9}$$

2.2.2 Plus Shaped Light Lattice

The plus shaped light lattice is made up of five point sources in Fourier space, oriented such that they form a plus. To account for polarisations, this specific example chooses the central source polarised at 45°, the vertically displaced sources horizontally polarised and the horizontally displaced sources vertically polarised. This is the lattice shown in Figure 10(b). The sources are assumed to each have the same amplitude, A. The resulting electric field is shown in Equation 10.

$$\vec{E}(\vec{r}) = A \begin{pmatrix} \frac{1}{\sqrt{2}} + e^{i2\pi k_x y} + e^{-i2\pi k_x y} \\ \frac{1}{\sqrt{2}} + e^{i2\pi k_y x} + e^{-i2\pi k_y x} \end{pmatrix} \tag{10}$$

The Fourier transformed electric field components are shown in Equations 11 and 12.

$$\hat{E}_x(\vec{k}) = \frac{A}{\sqrt{2}}\delta(\vec{k}) + A\delta(\vec{k} - \vec{k}_x) + A\delta(\vec{k} + \vec{k}_x) \tag{11}$$

$$\hat{E}_y(\vec{k}) = \frac{A}{\sqrt{2}}\delta(\vec{k}) + A\delta(\vec{k} - \vec{k}_y) + A\delta(\vec{k} + \vec{k}_y) \tag{12}$$

The illumination pattern in the Fourier domain is then determined by computing the 2D autocorrelation as shown in Equation 13.

$$\hat{h}(\vec{q}) = (\hat{E}_x \star \hat{E}_x)(\vec{q}) + (\hat{E}_y \star \hat{E}_y)(\vec{q}) \tag{13}$$

Computing this autocorrelation of delta peaks results in a new set of delta peaks located at the vector differences $\Delta\vec{k}$ between sources. By evaluating these differences, there are 3 distinct types of interactions:

1) Difference of $\vec{0}$: This case concerns the self-interactions of centre-centre, top-top, bottom-bottom, right-right and left-left. This term is often denoted as the DC component.

2) Difference of $\pm\vec{k}_x$ and $\pm\vec{k}_y$. This case concerns the central source interacting with the 4 outer sources.

3) Difference of $\pm 2\vec{k}_x$ and $\pm 2\vec{k}_y$. This case concerns the top-bottom and right-left interactions.

Because the horizontally displaced and vertically displaced sources have orthogonal linear polarisations, there are no interactions for vector differences of $\pm\vec{k}_x \pm \vec{k}_y$. The resulting illumination pattern representation is shown in Equation 14.

$$\hat{h}(\vec{q}) = \sum_m \hat{a}_m \delta(\vec{q} - \vec{k}_m) \tag{14}$$

In this equation $\vec{k}_m \in \{0, \pm\vec{k}_y, \pm 2\vec{k}_y, \pm\vec{k}_x, \pm 2\vec{k}_x\}$ and the amplitudes $\hat{a}_m$ are determined by multiplying the amplitudes of the interfering beams at those points. The exact amplitudes for this case are given in Table 3. This mathematically results in a distinct plus shaped geometry in Fourier space, as shown in Figure 13. Note that Figure 13 is displaying the 3D case, thus not all interactions are directly shown in this axial slice, which will be explained in more detail in a later section.

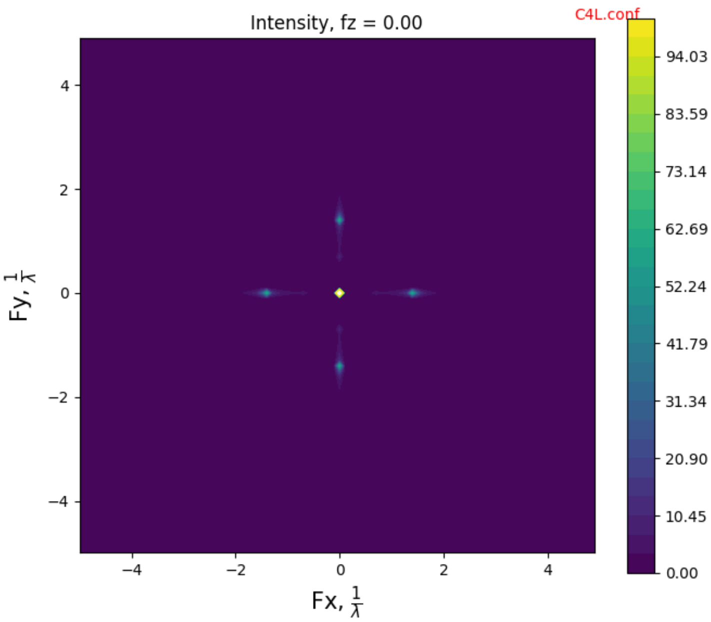

Figure 13: The resulting illumination pattern in 3D Fourier space of the plus shaped light lattice, viewed at an axial slice of $f_z = 0.00$. Figure generated with code from [26], where $|\vec{k}_x| = |\vec{k}_y| = 0.75$. The colour bar on the right indicates the intensity of the illumination pattern at a given point, using $A = 3$.

Finally, these illumination patterns are translated in space between images, hence to get the exact illumination pattern based on the shift, this must be included in this formula. To add the phase shift, $\vec{u}_p$, note that $h_{\text{shifted}}(\vec{r}) = h(\vec{r} - \vec{u}_p)$ thus $\mathcal{F}\{h(\vec{r} - \vec{u}_p)\} = \hat{h}(\vec{q})e^{-2\pi i\vec{q}\cdot\vec{u}_p}$. Including the phase shift gives the final representation shown in Equation 15.

$$\hat{h}(\vec{q}) = \sum_m \hat{a}_m e^{-2\pi i\vec{k}_m\cdot\vec{u}_p}\delta(\vec{q} - \vec{k}_m) \tag{15}$$

Equation 15 can be applied for the general case, where only the parameter sets for $\hat{a}_m$ and $\vec{k}_m$ must be found. A quick notable comparison is for the same case, except with the central source having a circular polarisation. The key difference in this case is that the polarisation of the central source goes from $\frac{1}{\sqrt{2}}(\hat{x} + \hat{y})$ to $\frac{1}{\sqrt{2}}(\hat{x} + i\hat{y})$.

With this the horizontally polarised components are unchanged, however the vertically polarised components now get complex amplitudes. This will change the amplitudes found from the autocorrelation between the centre and the horizontal sources such that $\hat{a} = A \cdot (\frac{A}{\sqrt{2}} i)^* = -i\frac{A^2}{\sqrt{2}}$ is now complex valued. This can be observed in Table 4. The spatial representation of this illumination pattern is displayed in Figure 14. The resulting pattern is a periodic plus shaped pattern.

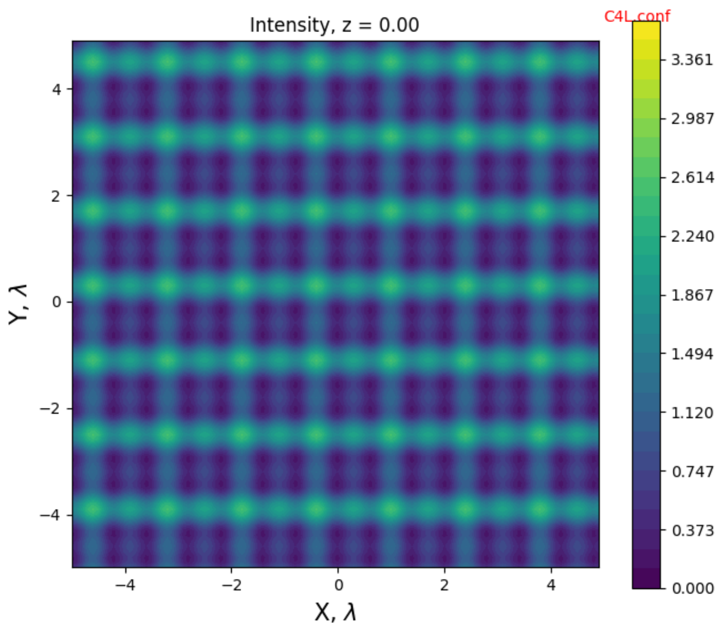

Figure 14: The resulting illumination pattern in the 3D spatial space for a plus shaped light lattice. Observe the periodic plus shaped pattern throughout space for this axial slice at $z = 0.00$. Figure generated with code from [26], where $|\vec{k}_x| = |\vec{k}_y| = 0.75$. The colour bar on the right indicates the intensity of the illumination pattern at a given point, using $A = 3$.

2.2.3 More General Lattice Shapes

More general forms can be created in a similar manner. It gets more interesting when looking at shapes that are not rectangular. Consider now a hexagonal pattern, as seen in Figure 10(d). For this, defining the position of the sources in k-space requires new basis vectors to be defined, specifically, using two basis vectors that are separated by 60^o. This can be represented by the basis vectors: $\vec{b}_1 = k_l(1, 0)$ and $\vec{b}_2 = k_l(\frac{1}{2}, \frac{\sqrt{3}}{2})$.

With this definition, the hexagonal lattice is defined by source terms located at: the central source $\vec{k}_0 = (0, 0)$ and the six surrounding sources are at locations

$\vec{k}_1, .., \vec{k}_6$ which take values $\vec{b}_1, -\vec{b}_1, \vec{b}_2, -\vec{b}_2, \vec{b}_2 - \vec{b}_1, -\vec{b}_2 + \vec{b}_1$. The illumination pattern can then be derived by computing the autocorrelation to be a sum of deltas at the points of interaction. See Table 5 in the appendix for the result. The visualisation for this specific light lattice in 3D Fourier space is shown in Figure 15 and Figure 16. In Figure 15 the self interactions can also be observed in the chosen axial slice, whilst in Figure 16 only the outer ring of sources can be observed.

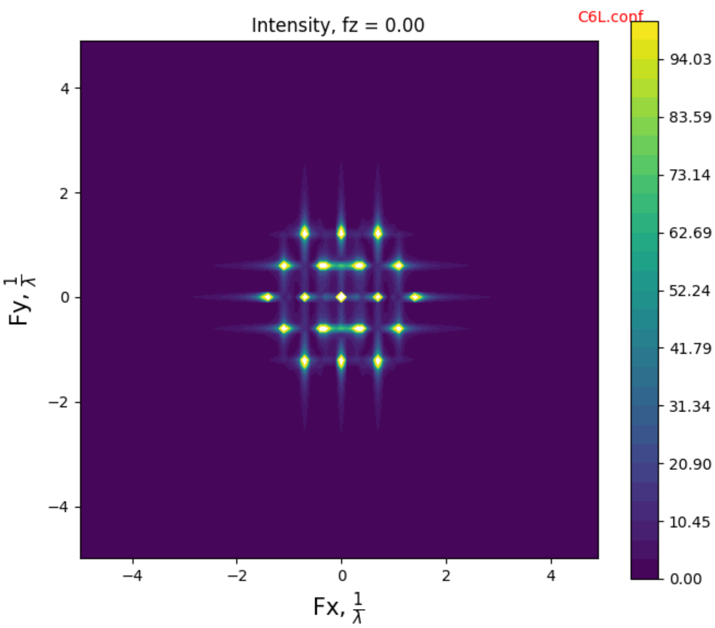

Figure 15: The resulting illumination pattern in 3D Fourier space of the hexagonal shaped light lattice, viewed at an axial slice of $f_z = 0.00$. Figure generated with code from [26], where $|\vec{k}_x| = |\vec{k}_y| = 0.75$. The colour bar on the right indicates the intensity of the illumination pattern at a given point, using $A = 3$.

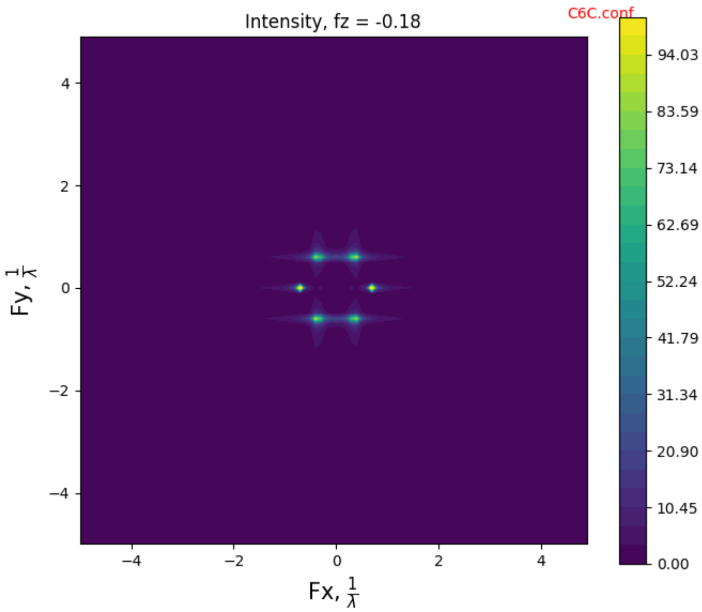

Figure 16: The resulting illumination pattern in 3D Fourier space of the hexagonal shaped light lattice, viewed at an axial slice of $f_z = -0.18$. Figure generated with code from [26], where $|\vec{k}_x| = |\vec{k}_y| = 0.75$. The colour bar on the right indicates the intensity of the illumination pattern at a given point, using $A = 3$.

2.2.4 Expansion to 3D

For 3D SIM the z component must also be considered. This is because moving the light lattice further from the sample results in the off-centre sources hitting the sample at different positions in the back focal plane (BFP), hence the positions of the sources in k-space are dependent on z. To expand the hexagonal pattern to the 3D case, the z dimension is added by taking note of the dispersion relation that requires $|\vec{k}|^2 = k^2$. The central beam therefore is now represented by $\vec{k}_0 = (0, 0, k)$. The six outer sources get a θ dependency from the incident angle, but also a ϕ dependency from azimuthal angle. For the hexagonal pattern these azimuthal angles are $\phi_m = (m-1)60°$ for $m = 1, \ldots, 6$. Consequently, the newly determined $\vec{k}$ vectors are $\vec{k}_m = (k \sin\theta \cos\phi_m, k \sin\theta \sin\phi_m, k \cos\theta)$.

These newly defined $\vec{k}$ vectors can then be used to determine the illumination pattern by computing the vector differences $\Delta\vec{k}_{i,j} = \vec{k}_i - \vec{k}_j$. The resulting illumination pattern expression is shown in Equation 16.

$$\hat{h}(\vec{q}) = \sum_{i=0}^{6} \sum_{j=0}^{6} \hat{a}_{i,j} \delta(\vec{q} - \Delta\vec{k}_{i,j}) \tag{16}$$

Note that for this the polarisation is not considered, adding polarisations can reduce the number of interactions to be computed. For this light lattice using circularly polarised plane waves, the representation of the illumination pattern in the spatial domain is shown in Figure 17. Here, the repeating hexagonal shape can be observed and that the intensity of illumination pattern around the hexagonal will vary, depending on the axial slice chosen. This can be seen when comparing Figures 17 and 18.

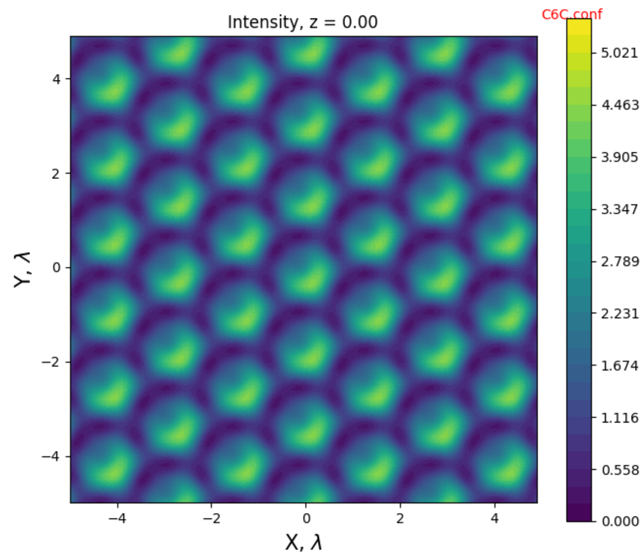

Figure 17: The resulting illumination pattern in the 3D spatial space for a hexagonal light lattice with all beams circularly polarised. Observe the greater intensity in the bottom half of the hexagon for the axial slice at $z = 0.00$. Figure generated with code from [26], where $|\vec{k}_x| = |\vec{k}_y| = 0.75$. The colour bar on the right indicates the intensity of the illumination pattern at a given point, using $A = 3$.

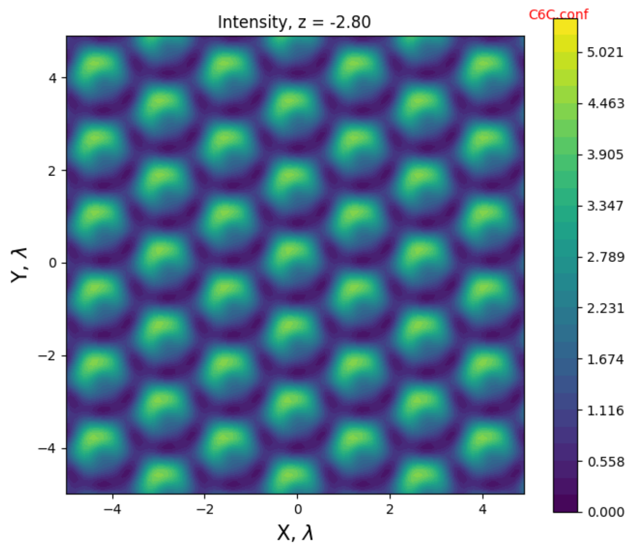

Figure 18: The resulting illumination pattern in the 3D spatial space for a hexagonal light lattice with all beams circularly polarised. Observe the greater intensity in the top half of the hexagon for the axial slice at $z = 0.00$. Figure generated with code from [26], where $|\vec{k}_x| = |\vec{k}_y| = 0.75$. The colour bar on the right indicates the intensity of the illumination pattern at a given point, using $A = 3$.

2.3 The Image

The image representation (in the spatial domain) can be found using Equation 17.

$$I(\vec{r'}) = \int g(\vec{r'} - \vec{r})h(\vec{r})f(\vec{r})dV \tag{17}$$

This equation assumes continuity, however imaging works with discrete pixels, hence a summation would be a more accurate depiction of the problem. The summation variant is shown in Equation 18. [19]

$$I_k^p = \sum_{l=1}^{N} g_{kl} h_l^p f_l \tag{18}$$

In Equation 18, f is the pixel representation of the object, h is the determined illumination pattern that hits the sample, g is the point spread function and I is the expected image. In this equation, the subscript k represents the discrete pixel indices of the final formed image and the superscript p indicates the specific translation (phase shift). N denotes the total number of pixels. To be able to determine the image formed, it is important to understand what the point spread function (PSF) looks like, which will be explored in this section.

2.3.1 Point Spread Functions (2D)

The point spread function (PSF) describes how an imaging system reacts to a point source of light. The light that is seen through a circular aperture in microscope lenses will form a pattern of circular disks, more precisely Airy disks, therefore in 2D the PSFs are the set of Airy disks. These Airy disks can be expressed by Equation 19 [19].

$$PSF(\vec{r}) = \left(\frac{2J_1(2\pi|\vec{r}|NA/\lambda)}{2\pi|\vec{r}|NA/\lambda} \right)^2 \tag{19}$$

In this equation, J_1 is the first order Bessel function, $|\vec{r}|$ the distance from the centre of the light source, NA the numerical aperture of the lens and λ the wavelength of the light used. A visualisation of the typical shape of an Airy disk is shown in Figure 19.

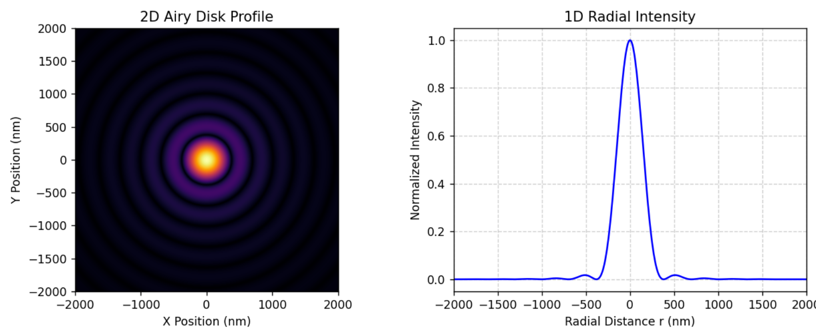

Figure 19: The left figure shows the typical intensity profile of an Airy disk for light entering the circular aperture of a lens in 2D. The figure on the right displays a plot of the normalised intensity as a function of the radial distance. This figure is made using the values NA = 0.8 and λ = 500 nm.

2.3.2 Optical Transfer Functions (2D)

The optical transfer function (OTF) is the Fourier transform of the PSF. The OTF is the result of the autocorrelation computed on the pupil function. For an optical microscope with a circular aperture, the pupil function is a flat open (circular) disk, therefore the OTF is the autocorrelation of the flat circular aperture. This autocorrelation and the Fourier transformation of the 2D Airy disks PSFs become what looks like a cone shape, as shown in Figure 20. The analytical expression for the OTF is known from the literature [27] and is given in Equation 20, with ρ the normalised spatial frequency spacing between adjacent sources, $\rho = \frac{k}{k_{\text{cutoff}}} = \frac{k}{\frac{2\text{NA}}{\lambda}}$.

$$g(\rho) = \begin{cases} \frac{2}{\pi}\left(\arccos(\rho) - \rho\sqrt{1-\rho^2}\right) & \text{for } 0 \le \rho \le 1 \\ 0 & \text{for } \rho > 1 \end{cases} \tag{20}$$

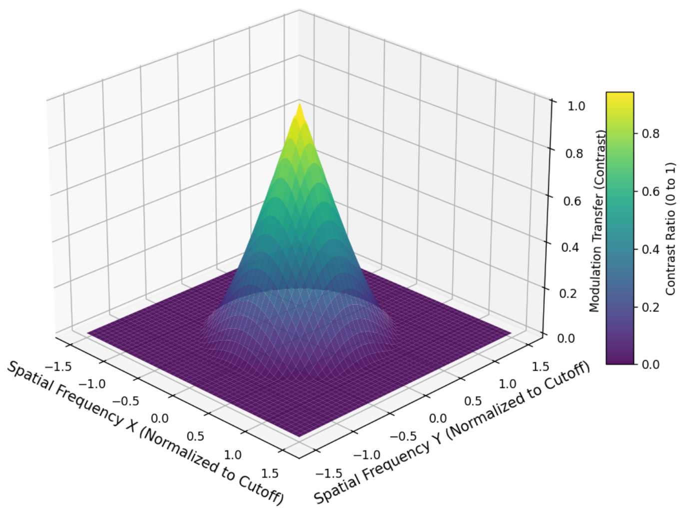

Figure 20: The optical transfer function in 2D for a circular aperture lens.

2.3.3 PSF and OTF (3D)

To get the resulting PSF in 3D it is not as simple as having layers of Airy disks for different axial slices. Instead, the resulting PSF is smeared along the optical axis as well as being spread out laterally, rather than forming a complete sphere.

In 3D the computed OTF takes the shape of a torus with the middle filled in. This torus is formed by a missing cone of light as a consequence of missing information at the centre along the q_z axis [28]. This missing double cone can be mathematically illustrated by analysing the conservation of energy. For a point source of light being imaged, when the light propagates along the optical axis, z, it forms the 3D PSF. Along any axial slice in the xy-plane the light will be spread out differently, however the total amount of energy will be conserved. This is shown in Equation 21.

$$\iint \mathrm{PSF}(x, y, z)dxdy = C \tag{21}$$

The OTF, $g(q_x, q_y, q_z)$, is the 3D Fourier transform of the PSF, given in Equation 22.

$$g(q_x, q_y, q_z) = \iiint \mathrm{PSF}(x, y, z)e^{-i2\pi(q_x x + q_y y + q_z z)}dxdydz \tag{22}$$

To analyse the 3D OTF along the optical axis, q_z, we assess the OTF at $g(0, 0, q_z)$ shown in Equation 23.

$$g(0, 0, q_z) = \int \left(\iint \mathrm{PSF}(x, y, z)dxdy \right) e^{-i2\pi q_z z}dz = \int Ce^{-i2\pi q_z z}dz \tag{23}$$

The final integral of the exponential in Equation 23 is the formal definition of the Dirac delta function, hence giving the result in Equation 24.

$$g(0,0,q_z) = C\delta(q_z) \tag{24}$$

In the Fourier domain this means that along the q_z axis that the transfer function is a delta peak and drops to 0 apart from at the origin. A visualisation of this OTF seen from a vertical slice is shown in Figure 21. This figure illustrates how this OTF axial slice takes the shape of a bow tie, where vertically there is a missing cone of light. This figure is plotted using the derived analytical 3D OTF formula from the literature for a circular, unaberrated pupil, shown in Equation 25 [29].

$$g(l,s) = \frac{2}{l}\mathrm{Re}\left[\sqrt{1-\left(\frac{|s|}{l}+\frac{l}{2}\right)^2}\right] \tag{25}$$

In Equation 25, $g(l,s)$ represents the 3D OTF, l is the normalised lateral spatial frequency magnitude, $\sqrt{q_x^2+q_y^2}$, and s is the normalised axial spatial frequency, q_z. Figure 21 is plotted at the lateral slice $q_y = 0$.

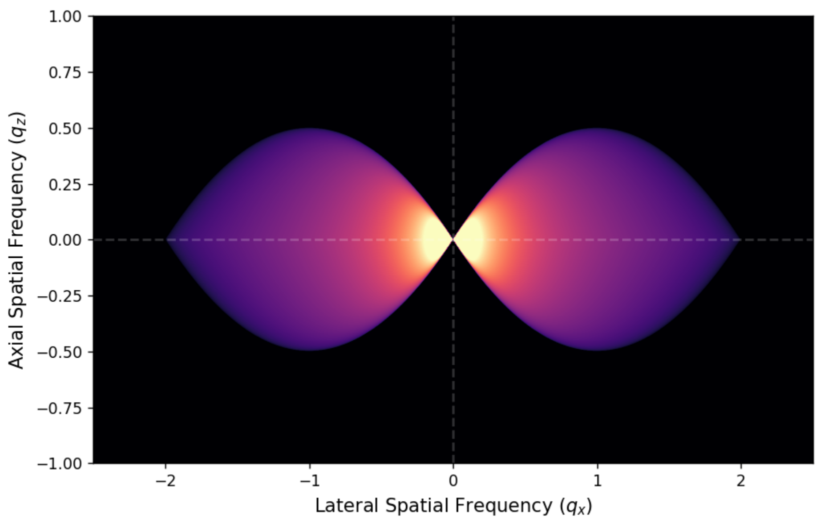

Figure 21: The 3D optical transfer function for a circular aperture lens as seen in the q_xq_z plane. This shows that along the optical axis q_z the OTF drops to 0, except at the origin, and that the general shape is that of a bow tie, with a lateral cutoff around $q_x = \pm 2.0$. The full OTF in 3D takes the shape of a torus that is filled in the centre.

In standard microscopy this missing double cone shape represents lost information in the image. 3D SIM is able to provide a solution for this issue by shifting the frequencies over to cover these gaps in Fourier space. By imaging multiple shifts, merging them and then finally separating them you are then able to reconstruct the image with all the information observable.

2.3.4 SIM Reconstruction Algorithm

The primary goal of the SIM reconstruction algorithm is to calculate an estimate, e_k, that best approximates the true object being observed, f_k, from the sequence of noisy images I_k^p. This problem is formulated as a regularised Least Squares problem, the derivation for which will be shown in this section [19]. The cost function to be minimised is given in Equation 26.

$$E = \frac{1}{2}\sum_{p=1}^{M}\sum_{k=1}^{N}\left(I_k^p - \sum_{l=1}^{N} g_{kl}h_l^p e_l\right)^2 + \frac{1}{2}\sum_{k,l=1}^{N} w_{kl}e_k e_l \tag{26}$$

In Equation 26, the summation bound N represents the total number of pixels in the image and M represents the total number of images taken during image acquisition. The subscript indices, k and l, denote individual pixels in the spatial domain and the superscript p indicates the specific phase shift considered. The key terms involved are the observed image at pixel k for phase shift p, I_k^p; the OTF function evaluated at vector difference between pixels k and l, g_{kl}; the illumination pattern at pixel l for phase shift p, h_l^p; the regularisation filter function, w_{kl} and the variables to be estimated, e_k. The first term in this expression represents the data fidelity, measuring the squared difference between the experimentally observed images and the theoretically expected images. The second term acts as a regularisation penalty to penalise noise and to make the solution stable. The regularisation filter taken for this problem is given in Equation 27.

$$w_{kl} = \frac{1}{N}\sum_{j=1}^{N}\hat{w}_j e^{2\pi i \vec{q}_j\cdot(\vec{r}_k - \vec{r}_l)} \tag{27}$$

In Equation 27, $\hat{w}_j$ is the regularisation filter kernel and $\vec{r}_k$, $\vec{r}_l$ are the pixel locations in real space for indices k and l. This kernel is usually taken to be constant across the frequency spectrum, leading to $\hat{w}_j = w$, $w_{kl} = w\delta_{kl}$.

To minimise the cost function in Equation 26, the partial derivatives with respect to the estimates e_j are set to 0, as shown in Equation 28.

$$\frac{\partial E}{\partial e_j} = -\sum_{p=1}^{M}\sum_{k=1}^{N} g_{kj}h_j^p\left(I_k^p - \sum_{l=1}^{N} g_{kl}h_l^p e_l\right) + \sum_{k=1}^{N} w_{jk}e_k = 0 \tag{28}$$

This can be written as a set of linear equations as given in Equation 29.

$$\sum_{l=1}^{N} T_{jl}e_l = b_j \tag{29}$$

In Equation 29, the different components are defined as shown in Equations 30 to 32.

$$b_j = \sum_{p=1}^{M}\sum_{k=1}^{N} h_j^p g_{kj} I_k^p \tag{30}$$

$$T_{jl} = w_{jl} + (g^T g)_{jl} C_{jl} \tag{31}$$

$$C_{jl} = \sum_{p=1}^{M} h_j^p h_l^p \tag{32}$$

This process gives the solutions to the regularised Least Squares problem given in Equation 33.

$$e_l = \sum_{j=1}^{N} T_{lj}^{-1} b_j \tag{33}$$

The overall reconstruction steps can be summarised by the following steps:
(1) The set of images I_k^n taken for all translations is Fourier transformed to $\hat{I}_k^n$.
(2) The distinct spatial frequency bands are untangled and extracted by applying appropriate linear combinations across the translated images.
(3) The spatial frequency bands are low-pass filtered. This project will use a custom derived low-pass filter, outlined in the later sections.
(4) The low-pass filtered frequency bands are shifted back to their correct positions in Fourier space with the centre band spatial frequency $\vec{k}_m$.
(5) A weighted sum over the shifted, low-pass filtered, spatial frequency bands is taken, which yields a weighted-sum-of-bands image.
(6) The weighted-sum-of-bands image is filtered with a linear filter to give the final image.

2.4 Spectral Signal to Noise Ratio

During image reconstruction, consideration must be taken of the accumulated noise. In this section the spectral signal-to-noise ratio (SSNR) will be derived for the general situation that will be considered throughout this project.

2.4.1 Imaging Noise in General

All images are corrupted by some form of noise. The extent of this is quantified by the signal-to-noise ratio. For general imaging purposes, a few key types of noise are encountered: shot noise, readout noise and thermal noise [30]. In this project the primary form of noise accounted for is the shot noise.

Shot noise is noise that arises during photon detection. Light is described as photons which are quanta of energy. For typical light microscopes, the number of detected photons in a fixed time interval is a stochastic variable that follows a Poisson distribution [31]. The noise as a consequence of the variation in the intensity of photons detected is called shot noise.

During the mathematical reconstruction in SIM the shot noise is amplified. The SIM algorithm works by extracting the high spatial frequency information and shifting it back to the correct position in frequency space. To restore proper contrast these high-frequency signals must be amplified during the filtering process. This inevitably will also amplify the shot noise alongside the signal.

2.4.2 Noise in SIM

SIM does not exclusively suffer from stochastic noise, but also an artificial structured noise artifact that arises due to the algorithm used. Specifically, the chosen filters dictates the structured artifact. In this subsection we will explicitly derive the relation for the SSNR [19], but for a custom chosen low pass filter.

During lattice SIM a total of M_t images are taken, determined by the number of translations of the illumination pattern used. Thus, $n = 1, \dots M_t$. In physical imaging the expected image (μ) differs from the real image (I) by some noise term (s) giving rise to Equation 34.

$$I_j^n = \mu_j^n + s_j^n \tag{34}$$

In Equation 34, the j subscript indexes refer to the individual pixels in real space. The superscript n denotes the translation. The noise, s, is typically the Poisson shot noise as discussed, which has zero mean for this application, shown in Equation 35.

$$E[s_j^n] = E[I_j^n] - \mu_j^n = \mu_j^n - \mu_j^n = 0 \tag{35}$$

To determine the power of the noise the correlation must be computed, $\langle s_j^n s_{j'}^{n'} \rangle$. The result of this correlation is depicted in Equation 36.

$$\langle s_j^n s_{j'}^{n'} \rangle = \delta_{nn'} \delta_{jj'} \mu_j^n \tag{36}$$

The noise correlation in Fourier space becomes as shown in Equation 37, where the new subscripts l and l' are the Fourier space indices.

$$\langle \hat{s}_l^n \hat{s}_{l'}^{n'} \rangle = \sum_{j=1}^{N} \sum_{j'=1}^{N} \langle s_j^n s_{j'}^{n'} \rangle e^{-2\pi i (\vec{q}_l \cdot \vec{r}_j - \vec{q}_{l'} \cdot \vec{r}_{j'})} = \hat{\mu}(\vec{q_l} - \vec{q_{l'}}) \tag{37}$$

During reconstruction more noise is accumulated. To isolate the structural information, the different spatial frequency bands are extracted by taking a linear combination of the translated images. For the mth-order band this is denoted by the expression in Equation 38.

$$\hat{J}_l^m = \sum_{n=1}^{M_t} e^{-2\pi i \vec{k}_m \cdot \vec{u}_n} \hat{I}_l^n \tag{38}$$

In Equation 38, $\vec{u}_n$ is the phase shift. The noise components then also combine, which yields a new cross-correlation between orders as determined in Equation 39.

$$\langle \delta \hat{J}_l^m \delta \hat{J}_{l'}^{m'*} \rangle = M_t \hat{a}_{m-m'} \hat{g}(\vec{q}_l - \vec{q}_{l'}) \hat{f}(\vec{q}_l - \vec{q}_{l'} + \vec{k}_m - \vec{k}_{m'}) \tag{39}$$

The next step in this derivation is unique to this project. The typical derivation assumes filtering with the conjugate of the microscope OTF [19], however this project will utilise a custom filter, $\hat{K}_m(\vec{q}_l)$. This filter yields the filtered bands $\hat{L}_l^m = \hat{K}_m(\vec{q}_l)\hat{J}_l^m$, where the final noise correlation is then updated to the expression in Equation 40.

$$\langle \delta \hat{L}_l^m \delta \hat{L}_{l'}^{m'*} \rangle = M_t \hat{a}_{m-m'} \hat{K}_m(\vec{q}_l) \hat{K}_{m'}^*(\vec{q}_{l'}) \hat{g}(\vec{q}_l - \vec{q}_{l'}) \hat{f}(\vec{q}_l - \vec{q}_{l'} + \vec{k}_m - \vec{k}_{m'}) \tag{40}$$

Next, these filtered bands are shifted in Fourier space to their correct centre frequencies producing bands $\hat{L}^m(\vec{q}_l - \vec{k}_m)$. These are recombined using a linear combination of weighted sums giving the intermediate images, shown in Equation 41.

$$\hat{b}_l = \sum_{m \in \mathbb{M}} \hat{c}_m \hat{L}^m(\vec{q}_l - \vec{k}_m) \tag{41}$$

From the intermediate images given by Equation 41 the critical components for the SSNR can be extracted. The signal transfer function, $\hat{D}_l$, indicates the final strength of the signal, which is affected by the custom filter and the chosen weights. This is derived by taking the expected value of Equation 41, shown in Equation 42.

$$\langle \hat{b}_l \rangle = \sum_{m \in \mathbb{N}} \hat{c}_m \left[M_t \hat{a}_m \hat{K}_m(\vec{q}_l - \vec{k}_m) \hat{g}(\vec{q}_l - \vec{k}_m) \hat{f}(\vec{q}_l) \right] \tag{42}$$

From Equation 42 the signal transfer function, $\hat{D}_l$, can be extracted by looking at the coefficient of the $\hat{f}(\vec{q}_l)$ term, shown in Equation 43.

$$\hat{D}_l = M_t \sum_{m \in \mathbb{M}} \hat{a}_m \hat{c}_m \hat{K}_m(\vec{q}_l - \vec{k}_m) \hat{g}(\vec{q}_l - \vec{k}_m) \tag{43}$$

The shot noise variance captures the cross-talk of the noise variance caused by overlapping the shifted bands. This is derived by taking the variance of $\hat{b}_l$, which gives the expression for the shot noise variance, $\hat{V}_l$, as shown in Equation 44.

$$\hat{V}_l = M_t \sum_{m,m' \in \mathbb{M}} \hat{c}_m \hat{c}_{m'}^* \hat{a}_{m-m'} \hat{K}_m(\vec{q}_l - \vec{k}_m) \hat{K}_{m'}^*(\vec{q}_l - \vec{k}_{m'}) \hat{g}(\vec{k}_{m'} - \vec{k}_m) \tag{44}$$

Finally, by dividing the squared signal power by the noise variance, the final SSNR expression is determined, shown in Equation 45.

$$SSNR_l = \frac{|\hat{D}_l|^2 |\hat{f}_l|^2}{\hat{V}_l \hat{f}_0} \tag{45}$$

2.4.3 Expressing the SSNR as an Inner Product

Observe that the signal transfer function, $\hat{D}_l$, and the shot noise, $\hat{V}_l$, in Equations 43 and 44 respectively, are given by summations for the same indexed m and m', and therefore these expressions can be written more compactly as inner products, as shown in Equations 46 and 47.

$$\hat{D}_l = M_t \langle x_l | y_l \rangle \tag{46}$$

$$\hat{V}_l = M_t \langle x_l | M | x_l \rangle \tag{47}$$

In these equations $|x_l\rangle$ and $|y_l\rangle$ are vectors defined by Equations 48 and 49, whilst M is a matrix given by Equation 50.

$$x_l = \hat{c}_m \hat{K}_m (\vec{q}_l - \vec{k}_m) \tag{48}$$

$$y_l = \hat{a}_m \hat{g}(\vec{q}_l - \vec{k}_m) \tag{49}$$

$$M_{m'm} = \hat{a}_{m-m'} \hat{g}(\vec{k}_{m'} - \vec{k}_m) \tag{50}$$

This work yields the more compact expression of the SSNR shown in Equation 51 which will be used in later sections.

$$SSNR_l = \frac{|\langle x_l | y_l \rangle|^2}{\langle x_l | M | x_l \rangle} \frac{|\hat{f}|^2}{\hat{f}_0} \tag{51}$$

3 Computing Analytical Bounds for SSNR

Working without the assumption that $M = I$ requires additional analytical or numerical steps to be taken in the reconstruction process. The convention $M = I$ makes the low-pass step an easy task since computation with the identity is simple. This section attempts to find analytical bounds for the SSNR.

3.1 Simplified Upper Bound

For physical imaging maximising the SSNR is desirable to produce the optimal image. The problem therefore lies in trying to maximise the expression in Equation 51. To simplify this problem, conventionally it is taken that the matrix M is the identity matrix. This simplification is justified since the OTF decreases in magnitude for larger difference indices, the matrix will consequently be dominated by the main diagonal, which can be approximated to the identity. This simplification yields the expression and inequality shown in Equation 52, which directly computes the upper bound.

$$\frac{|\langle x|y\rangle|^2}{\langle x|x\rangle} \leq \frac{\langle x|x\rangle \langle y|y\rangle}{\langle x|x\rangle} = \langle y|y\rangle \tag{52}$$

This is an interesting result, since this result claims that the maximal value occurs precisely when $x = y$, or put into context this means that the low-pass filter that should be chosen is actually precisely the OTF, which was directly assumed and used in the supplementary derivation [19]. Naturally, this is not a perfect conclusion and the assumption that $M = I$ will generally not be a good one. This project aims to assess to what extent this assumption can be improved upon by not assuming that $M = I$ and rather working directly with the exact M matrix. To work with the noise matrix directly, we require that M is invertible. This will be explored in the next section.

3.2 More Accurate Upper Bound

3.2.1 Showing that the Noise Matrix is Positive Definite

Before starting to compute the more accurate upper bound, it must be established that M is positive definite, since this will ensure that M is invertible. The property can be derived intuitively and rigorously.

From a physical standpoint, the term $\langle x| M |x\rangle$ represents the spectral variance of the shot noise. Because photon counts follows Poisson statistics [31], this variance will be strictly positive, provided that light is being shone to ensure the strict inequality. Consequently, $\langle x| M |x\rangle > 0$, thus indicating that M must be positive definite. This proof is the most intuitive and direct.

Alternatively, this can be argued from a more rigorous mathematical perspective. The matrix M is defined as $M_{m'm} = \hat{a}_{m-m'}\hat{g}(\vec{k}_{m'} - \vec{k}_m)$. Because

$h(\vec{r})$ is purely real valued, its Fourier transform will be conjugate symmetric, meaning that $\hat{a}_{-m} = \hat{a}^*_m$. Similarly, the PSF is also real valued, hence the OTF will also be conjugate symmetric, thus $\hat{g}(-\vec{k}) = \hat{g}(\vec{k})^*$. From this follows the relation in Equation 53, proving that M is Hermitian.

$$M^*_{m'm} = \Big(\hat{a}_{m-m'}\hat{g}(\vec{k}_{m'} - \vec{k}_m)\Big)^* = \hat{a}_{-(m-m')}\hat{g}(-(\vec{k}_{m'} - \vec{k}_m)) = M_{mm'} \quad (53)$$

The central diagonal elements are given by $M_{mm} = \hat{a}_0\hat{g}(0)$, where $\hat{a}_0$ is the strictly positive, real DC component (global average intensity) of the illumination pattern and $\hat{g}(0)$ is the strictly positive, real peak of the OTF at zero frequency. Assuming strict diagonal dominance ($|M_{ii}| > \sum_{j\neq i} |M_{ij}|$), the Gershgorin Circle Theorem [32] can be applied. This is a generally valid assumption, since the value of the OTF quickly drops off, hence the central diagonal is expected to be larger than the off-centre diagonals.

The Gershgorin Circle Theorem states that every eigenvalue of M must lie within a disc centred at M_{ii} with radius equal to the sum of the off-diagonal magnitudes. The assumption of strict dominant convergence ensures that each disc is strictly confined to the positive real space, therefore that all the eigenvalues of M are real and positive, directly proving that M is positive definite.

3.2.2 Exact Upper Bound

The previous upper bound derivation approximated M with the identity. However if we use the property that M is invertible this assumption is not necessary. Using that M is invertible, a more precise upper bound can be derived by the steps outlined in Equations 54 and 55. This yields the new optimality condition exactly when $x = M^{-1}y$.

$$\frac{|\langle x|y\rangle|^2}{\langle x|M|x\rangle} = \frac{|\left\langle M^{\frac{1}{2}}x \middle| M^{-\frac{1}{2}}y\right\rangle|^2}{\left\langle M^{\frac{1}{2}}x \middle| M^{\frac{1}{2}}x\right\rangle} \leq \frac{\left\langle M^{\frac{1}{2}}x \middle| M^{\frac{1}{2}}x\right\rangle \left\langle M^{-\frac{1}{2}}y \middle| M^{-\frac{1}{2}}y\right\rangle}{\left\langle M^{\frac{1}{2}}x \middle| M^{\frac{1}{2}}x\right\rangle} \quad (54)$$

$$\frac{|\langle x|y\rangle|^2}{\langle x|M|x\rangle} \leq \langle y|M^{-1}|y\rangle \quad (55)$$

4 Computing Bounds for the SSNR Ratio

To assess the impact of this refinement derived in the previous chapter, the difference between taking $x = M^{-1}y$ (optimal) and $x = y$ (suboptimal) must be evaluated. This can be expressed as a ratio as shown in Equations 56 and 57.

$$\frac{\text{Optimal SSNR}}{\text{Suboptimal SSNR}} = \frac{\langle y| M^{-1} |y\rangle}{\frac{|\langle y|y\rangle|^2}{\langle y|M|y\rangle}} = \frac{\langle y| M^{-1} |y\rangle \langle y| M |y\rangle}{| \langle y|y\rangle |^2} \tag{56}$$

$$\frac{\text{Suboptimal SSNR}}{\text{Optimal SSNR}} = \frac{| \langle y|y\rangle |^2}{\langle y| M^{-1} |y\rangle \langle y| M |y\rangle} \tag{57}$$

Within this chapter, the SSNR ratio will be determined for a selection of light lattices. For many of the commonly used light lattices computing the SSNR ratio analytically is difficult, therefore this chapter aims to find bounds for the SSNR ratio. This project aims to assess whether taking these additional steps will provide a quantifiable improvement to the SSNR ratio. The next section introduces useful theorems for determining these bounds.

4.1 Useful Properties and Theorems for the Noise Matrix

This project works extensively with the noise matrix, M, given by $M_{m'm} = \hat{a}_{m-m'}\hat{g}(\vec{k}_{m'} - \vec{k}_m)$. It is therefore helpful to introduce key theorems that can be applied to this matrix to bound the SSNR ratio.

4.1.1 Kantorovich Inequality

The Kantorovich Inequality [33] is a special case of the triangle inequality which bounds how much a positive definite matrix and its inverse can stretch a vector. The Kantorovich inequality states that for an $n \times n$ symmetric, positive definite matrix A with maximum and minimum eigenvalues λ_{max} and λ_{min}, and non-zero vector x, then the inequality shown in Equation 58 holds.

$$\frac{(x^T Ax)(x^T A^{-1}x)}{(x^T x)^2} \leq \frac{(\lambda_{max} + \lambda_{min})^2}{4\lambda_{max}\lambda_{min}} \tag{58}$$

The Kantorovich bound gives the worst case, hence for certain applications it may yield loose bounds. The bound value occurs when the vector x points along the eigenvectors corresponding to the extreme eigenvalues of A.

4.1.2 Toeplitz Matrices

A Toeplitz matrix [34] is a matrix where entries along diagonals is constant. An example of such a matrix is shown in Equation 59.

$$T = \begin{bmatrix} t_0 & t_{-1} & t_{-2} \\ t_1 & t_0 & t_{-1} \\ t_2 & t_1 & t_0 \end{bmatrix} \tag{59}$$

When the index spacing is constant between points in k-space, then the resulting noise matrix M will be Toeplitz. This can be seen when looking at the expression for M, where the values along each diagonal only depend on the index difference. If these are constant, then the resulting matrix will be Toeplitz.

4.1.3 Approximating Eigenvalues for Hermitian Toeplitz Matrix

Finding the eigenvalues of larger $n \times n$ matrices is challenging. Instead the symbol, or the generating function, of the matrix is considered. Define the diagonal entries of the matrix M as t_k, where since M is Hermitian $t_{-k} = \overline{t_k}$. These diagonals act as the Fourier coefficients for a real-valued periodic function $f(\theta)$ that is defined in Equation 60. Then the Bounding Theorem [34] states the result in Equation 61.

$$f(\theta) = \sum_{k=-(n-1)}^{n-1} t_k e^{ik\theta} \tag{60}$$

$$\min(f(\theta)) \leq \lambda_{\min} \leq \lambda \leq \lambda_{\max} \leq \max(f(\theta)) \tag{61}$$

This result also holds using the 2D generating function shown in Equation 62, where the elements t_k have been entered for the general case in SIM.

$$f(\theta_1, \theta_2) = \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} a_{p,q} g(p\vec{b}_1 + q\vec{b}_2) e^{i(p\theta_1 + q\theta_2)} \tag{62}$$

4.2 Bounding the SSNR Ratio

The bounds for the SSNR ratio can be easily determined using the Kantorovich inequality [33] described in the earlier section. Because M is a positive definite, Hermitian square matrix, this inequality can be directly applied to the ratio defined in Equation 56, giving the upper bound for this ratio expressed in Equation 63.

$$\frac{\text{Optimal SSNR}}{\text{Suboptimal SSNR}} = \frac{\langle y| M^{-1} |y\rangle \langle y| M |y\rangle}{| \langle y|y\rangle |^2} \leq \frac{(\lambda_{max} + \lambda_{min})^2}{4\lambda_{max}\lambda_{min}} \tag{63}$$

Taking the reciprocal then yields Equation 64, which provides a lower bound on how much worse the suboptimal filter choice would be. It is worth reiterating that the Kantorovich bound is a worst-case scenario, therefore often the true ratio may not be as poor.

$$\frac{\text{Suboptimal SSNR}}{\text{Optimal SSNR}} \geq \frac{4\lambda_{max}\lambda_{min}}{(\lambda_{max} + \lambda_{min})^2} \tag{64}$$

4.3 1D Lattice Structures

The most commonly used SIM configurations are the 1D light lattices, 2D conventional SIM and 3D conventional SIM. The SSNR ratios for these configurations will be determined in this section.

4.3.1 Example 2D Conventional SIM

In 2D conventional SIM the lattice is generated by two delta peaks in Fourier space. These two sources result in three Fourier orders in k-space once the autocorrelations have been computed. For this, the Fourier orders are labelled as $m \in \{-1, 0, 1\}, \quad a_{-1} = a_1, \quad a_1 g(k) = a_{-1} g(-k)$. Using the definition of $M_{m'm} = \hat{a}_{m-m'} \hat{g}(\vec{k}_{m'} - \vec{k}_m)$, the entries of M can be determined such that M takes the form shown in Equation 65.

$$M = \begin{pmatrix} a_0 \mathrm{g}(0) & a_1 g(k) & 0 \\ a_1 g(k) & a_0 \mathrm{g}(0) & a_1 g(k) \\ 0 & a_1 g(k) & a_0 \mathrm{g}(0) \end{pmatrix} \tag{65}$$

The eigenvalues can be determined and are shown in Equations 66 and 67.

$$\lambda_{max} = a_0 g(0) + \sqrt{2} a_1 g(k) \tag{66}$$

$$\lambda_{min} = a_0 g(0) - \sqrt{2} a_1 g(k) \tag{67}$$

Using these eigenvalues, the Kantorovich inequality bound is computed and shown in Equation 68.

$$\frac{\text{Suboptimal SSNR}}{\text{Optimal SSNR}} \geq 1 - 2 \left(\frac{a_1 g(k)}{a_0 g(0)} \right)^2 \tag{68}$$

This situation is sufficiently simple to allow a direct computation of the SSNR ratio by inverting M. An example calculation will be shown at the centre spatial frequency $\vec{q} = 0$. First we simplify M to the expression in Equation 69.

$$M = \begin{pmatrix} d & c & 0 \\ c & d & c \\ 0 & c & d \end{pmatrix} \tag{69}$$

In Equation 69, the simplified terms are given by $d = a_0 g(0)$ and $c = a_1 g(k)$. Using this, the inverse is determined as in Equation 70.

$$M^{-1} = \frac{1}{d(d^2 - 2c^2)} \begin{pmatrix} d^2 - c^2 & -cd & c^2 \\ -cd & d^2 & -cd \\ c^2 & -cd & d^2 - c^2 \end{pmatrix} \tag{70}$$

Evaluating each term in the bound then yields the results in Equations 71 to 73, where $y = (c, d, c)^T$.

$$|\langle y|y\rangle|^2 = d^4 + 4c^2d^2 + 4c^4 \tag{71}$$

$$\langle y|M|y\rangle = d(d^2 + 6c^2) \tag{72}$$

$$\langle y|M^{-1}|y\rangle = d \tag{73}$$

All of this work allows one to compute the direct value of the ratio, the expression is shown in Equation 74.

$$\frac{\text{Suboptimal SSNR}}{\text{Optimal SSNR}} = 1 - \frac{2c^2d^2 - 4c^4}{d^4 + 6c^2d^2} \tag{74}$$

The difference between the exact ratio and the bound for the ratio in this case is non-zero, suggesting that the penalty incurred by the suboptimal filter is less severe than the Kantorovic bound suggests for this lattice at zero spatial frequency. This difference is shown in Equation 75.

$$R_{\text{exact}} - R_{\text{K}} = \frac{16c^4}{d^2(d^2 + 6c^2)} \tag{75}$$

Since d corresponds to the central diagonal terms, it is expected to be larger than the off diagonal c. Therefore Equation 75 approximately has magnitude $\frac{c^4}{d^4}$. For small k values, c is larger since the OTF drops rapidly for increasing k. Therefore the difference is most noticeable at smaller k values.

4.3.2 Example 3D Conventional SIM

The next easiest lattice configuration is 2D SIM, with three point sources on a line, otherwise known as 3D conventional SIM. This lattice is generated by three delta peaks in Fourier space at positions $-k$, 0 and k, resulting in five Fourier orders at locations $m \in \{-2k, -k, 0, k, 2k\}$. This configuration nets the Toeplitz structure for the matrix M as seen in Equation 76.

$$M = \begin{pmatrix} a_0g(0) & a_1g(k) & a_2g(2k) & 0 & 0 \\ a_1g(k) & a_0g(0) & a_1g(k) & a_2g(2k) & 0 \\ a_2g(2k) & a_1g(k) & a_0g(0) & a_1g(k) & a_2g(2k) \\ 0 & a_2g(2k) & a_1g(k) & a_0g(0) & a_1g(k) \\ 0 & 0 & a_2g(2k) & a_1g(k) & a_0g(0) \end{pmatrix} \tag{76}$$

This matrix is already large dimensionally and despite the Hermitian, Toeplitz structure, direct computation of the eigenvalues is challenging. Instead, the bounding theorem will be utilised. By labelling the centre interaction $c_0 = a_0g(0)$, the first order $c_1 = c_{-1} = a_1g(k)$ and the second order $c_2 = c_{-2} = a_2g(2k)$, the bounding theorem nets Equation 77.

$$f(\theta) = c_0 + 2c_1\cos(\theta) + 2c_2\cos(2\theta) \tag{77}$$

This gives the eigenvalue estimates in Equations 78 and 79.

$$\lambda_{\text{max}} \approx f(0) = c_0 + 2c_1 + 2c_2 \tag{78}$$

$$\lambda_{\text{min}} \approx f(\pi) = c_0 - 2c_1 + 2c_2 \tag{79}$$

Using these eigenvalue estimaes, the Kantorovich inequality can be determined as in Equation 80.

$$\frac{\text{Suboptimal SSNR}}{\text{Optimal SSNR}} \geq 1 - \frac{4(a_1 g(k))^2}{(a_0 g(0) + 2a_2 g(2k))^2} \tag{80}$$

Inverting M to determine the exact SSNR ratio quickly becomes too complicated and hence a numerical approach for computing the exact SSNR ratio is favoured. This will be considered in the next section.

4.3.3 Numerical SSNR Ratio Bounds 2D and 3D Conventional SIM

The analytic inequalities for the SSNR ratios have been determined. To get an insight into the accuracy of these SSNR bounds, plots can be helpful. These plots will compare the inequality results found with a numerically determined SSNR ratio using python to perform the vector and matrix operations. For this, the values of the OTF and the amplitudes must be known. The 2D OTF for a circular aperture is known from the literature and an analytical expression can be used, which is shown in Equation 81, with ρ the normalised spatial frequency spacing between adjacent sources.

$$g(\rho) = \begin{cases} \frac{2}{\pi}\left(\arccos(\rho) - \rho\sqrt{1-\rho^2}\right) & \text{for } 0 \leq \rho \leq 1 \\ 0 & \text{for } \rho > 1 \end{cases} \tag{81}$$

We want to compare how using two point sources in 2D conventional SIM compares to using three point sources in 3D conventional SIM. For both of these cases, the exact SSNR ratio has been analytically and numerically determined using python. This is plotted as a function of the spatial frequencies alongside the Kantorovich bounds. This is used to assess how different the worse case situation is as compared to the exact case. This plot is shown in Figure 22.

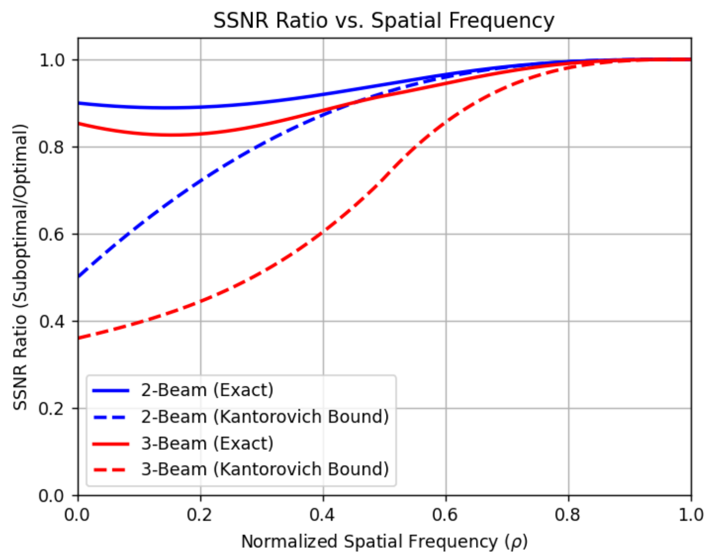

Figure 22: This figure shows a plot of the SSNR ratio for 2D conventional SIM and 3D conventional SIM as a function of the normalised spatial frequencies (ρ).

Figure 22 shows that in general the difference between the exact suboptimal and optimal SSNR is fairly small for both these two simple cases, and that taking the suboptimal choice $M = I$ becomes more impactful at smaller spatial frequencies or when adding more Fourier space delta peaks. The reason why the SSNR ratio dips for smaller spatial frequencies ρ, is because this represents cases where the Fourier orders lie closer together. This means that the vector difference between orders is smaller and hence the OTF evaluation is larger, resulting in more significant cross-talk entries into the noise matrix M.

Moreover, it is noticeable that the Kantorovich bound strongly under-estimates the SSNR ratio at the lower spatial frequencies. In this region the SSNR ratio considers a far worse scenario than what is physically realistic. To achieve the Kantorovich bound, the OTF must be made in a very specific way such that the OTF takes on the precise, discrete values. This is because the worst case only happens when the vector $y_m = a_m g(\vec{q}_l - \vec{k}_m)$ aligns perfectly along a combination of the eigenvectors of the minimum and maximum eigenvalues of M. Therefore to achieve this, the OTF $g(\vec{q}_l - \vec{k}_m)$ cannot be a smooth decaying curve. Since the PSF is the inverse Fourier transform of the OTF, the PSF would not be the typical Airy disk, but rather a spatial interference pattern. This is because the inverse transform of the discrete frequency peak OTF is a periodic wave, hence producing an interference pattern. This bound is therefore not to be taken as strictly achievable physically, but more as a measure to assess how bad the suboptimal filter could theoretically be.

4.4 2D Lattice Structures

In the prior section simple example lattice structures were discussed. Both of the discussed examples were 1D lattices, with sources aligned along the same axis. This section will explore more complex lattice shapes.

4.4.1 Example Plus Shaped Light Lattice

The 2D SIM with three beams lattice, seen in Figure 10(a), can be expanded by adding two new sources such that the lattice forms a plus shape. In k-space, the five active beams are located at $(0,0), (\pm k, 0)$ and $(0, \pm k)$. For this lattice the polarisation of the beams can be selectively chosen in different manners which can have an impact on the resulting noise matrix M. The considered cases are shown in Figure 10(b) and (c).

For the plus lattice with linearly polarised plane waves, there are 13 distinct Fourier orders, resulting in a 13×13 noise matrix M. The Fourier orders are shown in Table 3. The matrix representation for this lattice is given by Equation 82, where entries are given by $c(\vec{v}) = \hat{a}(\vec{v})\hat{g}(-\vec{v})$. The full M matrix with entries filled in is shown in the Appendix in Equation 102.

$$M_{\text{plus}} = \begin{bmatrix} c(\vec{q}_1 - \vec{q}_1) & c(\vec{q}_1 - \vec{q}_2) & \dots & c(\vec{q}_1 - \vec{q}_{13}) \\ c(\vec{q}_2 - \vec{q}_1) & c(\vec{q}_2 - \vec{q}_2) & \dots & c(\vec{q}_2 - \vec{q}_{13}) \\ \vdots & \vdots & \ddots & \vdots \\ c(\vec{q}_{13} - \vec{q}_1) & c(\vec{q}_{13} - \vec{q}_2) & \dots & c(\vec{q}_{13} - \vec{q}_{13}) \end{bmatrix} \tag{82}$$

The matrix form in Equation 82 may appear Toeplitz, however this is not the case. The jump from a line of points forming a 1D lattice to a 2D lattice results in the Toeplitz structure collapsing. Consider the index difference of -1. For this lattice, $\vec{q}_1 = (0,0,0)$, $\vec{q}_2 = (2,0,0)$, $\vec{q}_3 = (-2,0,0)$. Then on this diagonal, the entries become: $M_{1,2} = c(\vec{q}_1 - \vec{q}_2) = c((-2,0,0))$ and $M_{2,3} = c(\vec{q}_2 - \vec{q}_3) = c((4,0,0))$, which are not equal, despite having equal difference in index. Looking at the fully evaluated form in the Appendix in Equation 102 clearly displays the lack of Toeplitz structure. Since this matrix is both large dimensionally and lacks the Toeplitz structure, it is difficult to compute the eigenvalues analytically, hence another approach must be considered for this.

4.4.2 A Different Approach to the Plus Shaped Lattice

It was noted that the matrix in Equation 82 was not Toeplitz, but restoring the Toeplitz structure would be beneficial. This section considers a mathematical approach that reconstructs an acceptable Toeplitz type structure.

The problem is redefined as a 5×5 grid, containing 25 positions, with x and y coordinates from $x, y \in \{-2k, -k, 0, k, 2k\}$. The Fourier orders are known, and their coordinate locations are assigned the respective amplitudes, taken from either Table 3 or 4. The other coordinates in the grid where no Fourier order lies are zero padded, i.e. being assigned amplitude 0. Physically this represents an identical system, however mathematically it is slightly different.

With this mathematical approach, a Block-Toeplitz with Toeplitz-Blocks (BTTB) matrix is formed. This generates a 25×25 matrix, divided into smaller 5×5 blocks. This gives the macroscopic Toeplitz structure shown in Equation 83, where each entry is a 5×5 block, $T_{\Delta y}$.

$$M_{\text{plus}} = \begin{pmatrix} T_0 & T_{-1} & T_{-2} & T_{-3} & T_{-4} \\ T_1 & T_0 & T_{-1} & T_{-2} & T_{-3} \\ T_2 & T_1 & T_0 & T_{-1} & T_{-2} \\ T_3 & T_2 & T_1 & T_0 & T_{-1} \\ T_4 & T_3 & T_2 & T_1 & T_0 \end{pmatrix} \tag{83}$$

These 5×5 blocks, $T_{\Delta y}$, describe the physical interactions along the x-axis for a given shift in the y-axis. For example, for a vertical shift $\Delta y = 0$, the matrix T_0 is represented by the interaction of the central horizontal row with itself and is shown in Equation 84.

$$T_0 = \begin{pmatrix} a_{0,0}g(0,0) & a_{1,0}g(k,0) & a_{2,0}g(2k,0) & a_{3,0}g(3k,0) & a_{4,0}g(4k,0) \\ a_{-1,0}g(-k,0) & a_{0,0}g(0,0) & a_{1,0}g(k,0) & a_{2,0}g(2k,0) & a_{3,0}g(3k,0) \\ a_{-2,0}g(-2k,0) & a_{-1,0}g(-k,0) & a_{0,0}g(0,0) & a_{1,0}g(k,0) & a_{2,0}g(2k,0) \\ a_{-3,0}g(-3k,0) & a_{-2,0}g(-2k,0) & a_{-1,0}g(-k,0) & a_{0,0}g(0,0) & a_{1,0}g(k,0) \\ a_{-4,0}g(-4k,0) & a_{-3,0}g(-3k,0) & a_{-2,0}g(-2k,0) & a_{-1,0}g(-k,0) & a_{0,0}g(0,0) \end{pmatrix} \tag{84}$$

The amplitudes $a_{i,j}$ are taken from Table 3 or Table 4 depending on the respective polarisations. Further examples of the block matrices are shown in the Appendix in Equation 103 for T_1 and Equation 104 for T_2. This mathematical approach is useful since now the M_{plus} matrix is Toeplitz (Block-Toeplitz) and each submatrix T_i is likewise Toeplitz (Toeplitz-Blocks). The Toeplitz property will be helpful in approximating eigenvalues for much larger matrices since now the Bounding Theorem can be applied.

4.4.3 Applying the Bounding Theorem to the Plus Shape Lattice

With the newly derived BTTB M matrix the bounding theorem can be applied to eventually determine the Kantorovich bound. To begin, the entries are simplified to make computation easier. There are three types of physical interactions in this set up, namely the centre: $c_0 = (2 + 4r^2)\hat{g}(0)$, the first ring crosstalk: $c_1 = \hat{a}_{\pm 1,0}\hat{g}(k)$ and the 2nd ring crosstalk: $c_2 = -r^2\hat{g}(2k)$. The generating function then yields Equation 85, which has maximum and minimum given by Equations 86 and 87.

$$f(\theta_1, \theta_2) = c_0 + 2c_1(\cos\theta_1 + \cos\theta_2) + 2c_2(\cos 2\theta_1 + \cos 2\theta_2) \tag{85}$$

$$f(\theta_1, \theta_2)_{max} = f(0,0) = c_0 + 4c_1 + 4c_2 \tag{86}$$

$$f(\theta_1, \theta_2)_{min} = f(\pi, \pi) = c_0 - 4c_1 + 4c_2 \tag{87}$$

Note that the evaluation $f(\frac{\pi}{2}, \frac{\pi}{2}) = c_0 - 4c_2$ may seem smaller than the noted minimum in Equation 87, however we find that since the OTF falls off quickly, it means that $c_2 \ll c_1$, hence $-4c_1 + 4c_2 < -4c_2$. This means that the found minimum is indeed the minimum.

Using numerical techniques, approximate eigenvalues can also be determined for the entire 13×13 matrix. This is done to assess how much leeway the bounding theorem gives. This nets the maximum and minimum eigenvalues $\lambda_{max} \approx c_0 + 3c_1$ and $\lambda_{min} \approx c_0 - 3c_1$. For these two cases the Kantorovich inequality can be used giving the expression in Equation 88 for the bounding theorem result using the BTTB matrix and Equation 89 for the approximate Kantorovich bound for the 13×13 matrix.

$$R_{K,25} \geq 1 - \frac{16c_1^2}{(c_0 + 4c_2)^2} \tag{88}$$

$$R_{K,13} \geq 1 - \frac{9c_1^2}{c_0^2} \tag{89}$$

From this it can be observed that the true eigenvalues give a much tighter bound, hence being able to evaluate these will always be beneficial.

4.4.4 Approximating BTTB as Toeplitz

To assess how close to being Toeplitz the matrix in Equation 83 is, one can look at the diagonal entries. Along the main diagonal, the only entries are the values $c_{0,0}$, which is correct. Looking at the first diagonal for the sequence $M_{1,2}, M_{2,3}, M_{3,4} \dots M_{24,25}$, the Toeplitz structure is lost, since the possible entries are $c_{-1,0}$ and $c_{4,-1}$. These values are very different, with the former being non-zero and the latter being zero. Trying to approximate this with a Toeplitz matrix therefore does not work. Interestingly, making this assumption that they are equal leads to approximating $c_{-1,0}$, and other cross-talk terms, by 0 and hence results in $M = I$.

4.4.5 Numerical SSNR Ratio Bounds Plus Shaped Light Lattice

To assess the analytic bounds, plots are made comparing these to the numerically exact SSNR ratio. For this lattice, the 13×13 matrix was determined to not be Toeplitz, however using python the eigenvalues can still be determined numerically to determine the Kantorovich bound. In practice, this is not particularly useful since if using a numerical approach, one may as well directly use numerical inversion of the matrix to get the exact SSNR ratio. Regardless, this computational step was made to add another level of analysis to this lattice structure, producing the plot in Figure 23.

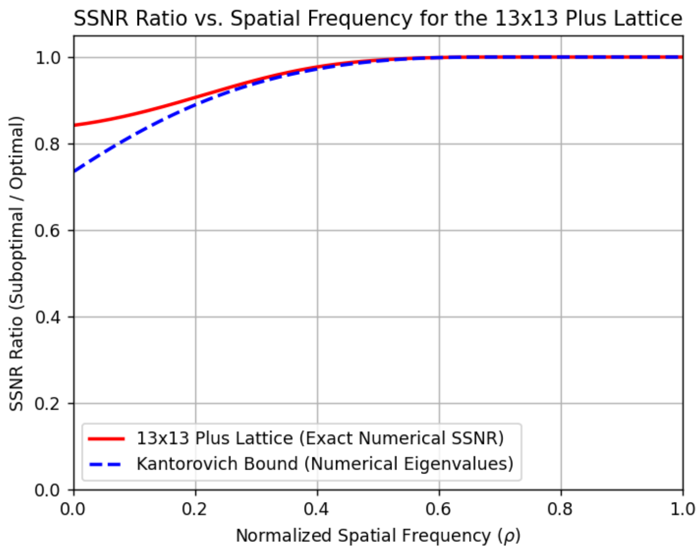

Figure 23: This figure shows a plot of the SSNR ratio for 3D SIM using a plus shaped light lattice with linearly polarised plane waves as a function of the normalised spatial frequencies. This figure is for the non-Toeplitz 13×13 matrix. Both the numerical exact and numerically bounded results are plotted.

Figure 23 suggests that the performance of the suboptimal filter again is generally great, decreasing in performance for the lower spatial frequencies but remaining above 0.8. Additionally, in this case, the Kantorovich bound netted a ratio that was lower as expected, however much less significant than in the previous cases with conventional SIM in Figure 22. The reason for this could be that the DC component, the central diagonal, is much stronger compared to conventional SIM, hence this dominates both the exact SSNR ratio and the maximum and minimum eigenvalue when numerically calculated.

Since the 13×13 matrix was not Toeplitz, a different mathematical approach was considered that produces a 25×25 BTTB matrix. This allows for approximating the SSNR using bounds rather than requiring numerical approximations. The result for this bound using the bounding theorem is plotted vs the numerically determined Kantorovich bound and the numerically exact SSNR ratio in Figure 24.

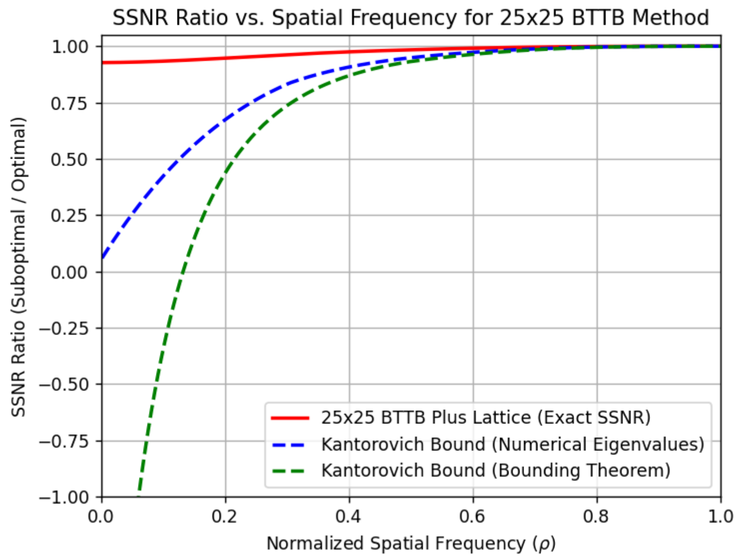

Figure 24: This figure shows a plot of the SSNR ratio for 3D SIM using a plus shaped lattice with linearly polarised plane waves as a function of the normalised spatial frequencies. This figure is for the BTTB 25×25 matrix. The result using the bounding theorem, the numerical exact and numerically bounded results are plotted.

Figure 24 shows that the exact SSNR curve remains the same, since mathematically they represent the same situations. Interestingly, the numerical Kantorovich bound is far worse. This may be since the central diagonal no longer has a clear dominance over the other diagonals. More noticeably, the approximated bound using the Bounding Theorem is so weak that it appears to drop to below zero, and even becomes very negative for small spatial frequencies. This result is unphysical, since the SSNR ratio should not be negative. The reason this value becomes negative can be observed from the Kantorovich bound, restated in Equation 90.

$$R_{K,25} \geq 1 - \frac{16c_1^2}{(c_0 + 4c_2)^2} \tag{90}$$

This equation becomes negative when the cross-talk from the first order, c_1, becomes so large that $4c_1 > c_0 + 4c_2$. At the smaller spatial frequencies where the bound becomes negative, the Fourier orders are crowded tightly together, which means that the cross-talk can become large enough to force the bound to become negative. This ultimately yields the conclusion that although enforcing this BTTB structure may be mathematically elegant, doing so for the purpose of determining analytical bounds for the SSNR ratio is not useful for analysis purposes. Rather, it is always best to invert the matrix M numerically to get an exact value for the SSNR ratio.

Finally, the circularly polarised plus lattice is considered. This results in a 17×17 matrix and produces the plot in Figure 25. This plot looks very similar to the result for linearly polarised plane waves as in Figure 23.

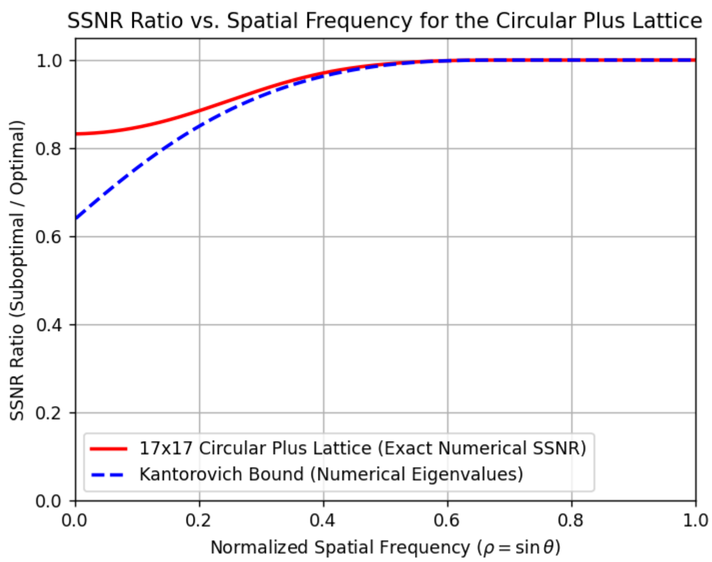

Figure 25: This figure shows a plot of the SSNR ratio for 3D SIM using a plus shaped lattice with circularly polarised plane waves as a function of the normalised spatial frequencies.

5 Evaluation of SSNR Ratios at Different Spatial Frequencies

5.1 Distinguishing Between Spatial Frequency Labels

In the mathematical analysis of SIM the term spatial frequency and its normalised value ρ are commonly used. However, this term is used to describe both the spatial frequency of the physical illumination pattern and the spatial frequency of the final reconstructed image. Depending on the context, the domain of the normalised spatial frequency changes.

In the earlier sections, the term spatial frequency was used to describe the physical illumination pattern, in k-space, that is projected onto the sample, ρ. This is the normalised spatial frequency spacing between adjacent sources, $\rho = \frac{k}{k_{\text{cutoff}}} = \frac{k}{\frac{2\text{NA}}{\lambda}}$. Because the light is passed through the objective lens to reach the sample, it is fundamentally constrained by the Abbe diffraction limit of that lens. Consequently, the spatial frequencies are limited by the lens cutoff ($\rho \leq 1$). In Figures 22 to 25 the spatial frequency label plotted on the x-axis therefore describes the effective pitch of the lattice.

In the following sections, the reconstructed image spatial frequency, $\vec{q}$, will be studied to assess the SSNR ratio in the reconstructed image. This refers to the spatial frequencies of the sample's structural details that are recovered in the super resolved output image. SIM allows the effective OTF to be expanded, recovering information up to twice the standard resolution limit. As a result, the cutoff spatial frequency is twice as large and the normalised spatial frequency axis expands to a maximum normalised value of $q = 2.0$.

5.2 Reconstructed Image SSNR Ratio 1D Lattices

The previous chapter evaluated the SSNR ratio for the different lattice structures at the spatial frequency $q = 0$, but for varying effective pitches $p = \frac{1}{\rho}$. This was required to compare the Kantorovich inequality performance, since this inequality is constant across spatial frequencies $\vec{q}$, however varies for the spatial frequencies ρ. This is because the Kantorovich inequality is only impacted by the noise matrix M, which has no dependence on $\vec{q}$. This section will examine how the SSNR ratio varies in the spatial frequency domain of the reconstructed image from $0 \leq q \leq 2$.

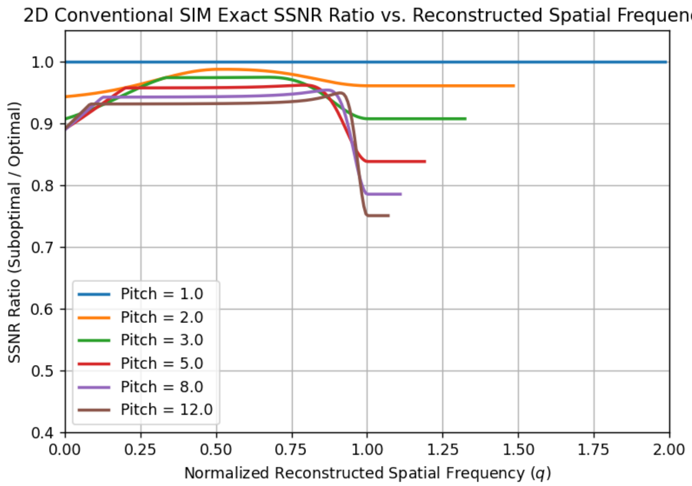

Figure 26: This figure shows how the SSNR ratio varies over the normalised spatial frequencies q for a range of different effective pitch values for 2D conventional SIM. These pitch values are the inverse of the normalised cutoff spatial frequencies, $p = \frac{1}{\rho}$.

For 2D conventional SIM the result is plotted in Figure 26 and the expected results as determined by Figure 22 can be read off, namely, the SSNR ratios at the spatial frequency $q = 0$. It can be observed that for increasing pitch (and therefore decreasing $\rho \to 0$), the SSNR ratio tends to a value just below 0.9 and for the pitch 1 ($\rho = 1$) the SSNR ratio is 1.

Figure 26 also clearly shows the amount of additional details that can be super resolved. Each curve abruptly stops at a specific $\vec{q}$ value, which corresponds to the maximum spatial frequency that can be observed. The value of the cutoff can be mathematically calculated. The maximum observable spatial frequency is determined by the Abbe diffraction limit, 1.0, plus the maximum frequency shift provided by the illumination pattern. For 2D conventional SIM, there are two Fourier delta peaks, giving a maximum Fourier order of $m = 1$. The largest frequency shift is therefore $m_{\max}\rho = \rho = \frac{1}{p}$. Therefore, for $p = 1$, it is expected that the maximum observable spatial frequency is $q_{max} = 1 + 1 = 2$, as shown in Figure 26. Doing a similar analysis yields that for $p = 2$, $q_{\max} = 1 + \frac{1}{2} = 1.5$ and for $p = 3$, $q_{max} = 1 + \frac{1}{3} = 1.33$. This result is also approximately shown in Figure 26.

The general shapes of the curves can also be understood through some physical intuition. For $p = 1$ the curve is a straight line with SSNR ratio 1.0, because for this case $M = I$ is perfectly valid. In this arrangement, the $m = \pm 1$ orders are shifted all the way to the edge of the cutoff limit. Since the airy disk patterns for each order are circles with radius 1.0, the orders do not overlap at all, resulting in zero cross-talk, validating that $M = I$ and the constant SSNR ratio.

For medium to large pitches the curves appear to all follow the same trend with the same features but with different values. In these cases the three Fourier orders become more tightly packed in the centre of the back focal plane (BFP). This results in large overlap of the OTFs and large amounts of cross-talk. This is felt strongly at the low spatial frequencies near $q = 0$ since this is near the centre of the pupil and hence the cross-talk is at its greatest. Moving away from the centre the overlap decreases and the cross-talk reduces, reducing the penalty of ignoring the cross-talk. The curves appear to flatten for $0.25 \leq q \leq 0.75$. In this range two effects are likely acting to oppose each other. The reducing OTF overlap pushes the SSNR ratio upwards, however, at the larger spatial frequencies the OTF falls off. As a result the structural information transferred by the optical system weakens. Therefore, for signals near the OTF boundary, it becomes more crucial to extract as much of the remaining detail as possible. Ignoring the cross-talk using $M = I$ therefore incurs an increased penalty closer to the OTF boundary.

Lastly, the curves have a sharp dip just before $q = 1$. At this point the central order $m = 0$ becomes very weak or stops passing through the lens entirely. Despite this, this central order still contributes to the cross-talk, causing a sharp penalty dip. Beyond $q = 1$ the SSNR ratio becomes constant. This is because at this point only a single Fourier order survives to provide information, hence no cross-talks can occur beyond this and the SSNR ratio only depends on the noise matrix rather than spatial frequencies $\vec{q}$.

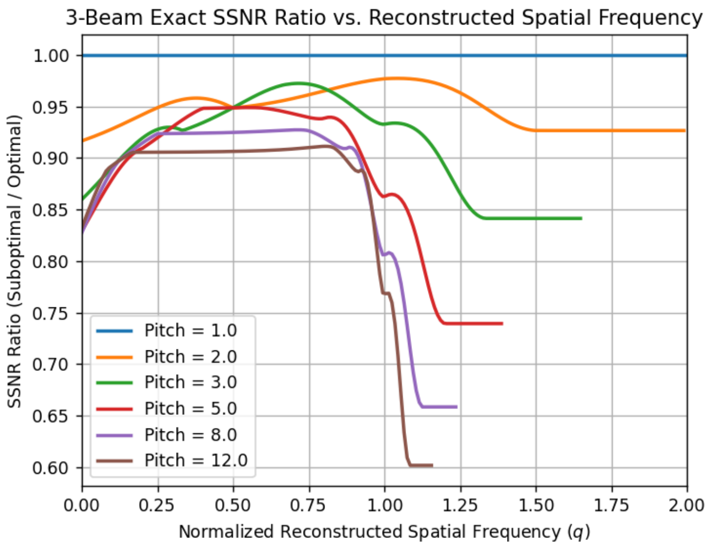

Figure 27: This figure shows how the SSNR ratio varies over the normalised spatial frequencies q for a range of different effective pitch values for the 3-beam lattice configuration. These pitch values are the normalised $p = \frac{1}{\rho}$.

Moving onto 3D conventional SIM, the result is shown in Figure 27 and a similar read off analysis can be performed to compare the results at $q = 0$. Beyond this spatial frequency similar effects and trends are observed as in the 2D conventional SIM case, only with additional complexities to the shape of the curves and the cutoff frequencies. This is due to the introduction of a third delta peak, resulting in more Fourier orders that can influence the shape of the curve. The most interesting feature is the cutoff frequency, which can be determined by $q_{\max} = 1 + m_{\max}\rho = 1 + \frac{2}{p}$. It is also interesting to note the differences between Figures 26 and 27. Notably, the cutoff frequency is increased when increasing the number of sources in the lattice, however this comes with the apparent drawback that in almost all cases the assumption $M = I$ is penalised more. This is an expected conclusion that more contributing Fourier orders result in potentially more information to be observed, however these will also have more cross-talk.

5.3 Reconstructed Image SSNR Ratio 2D Lattice

The same analysis is performed on the plus shaped lattice in this section. The shape of the curves for the plus lattice in Figure 28 is different from the 1D lattice types. The baseline SSNR ratio always remains quite large. This is likely because the DC component amplitude is large compared to the sidebands. This ensures that the noise matrix M is more diagonally dominant, resulting in minimal penalty for using the suboptimal filter. The cutoff frequency in this plot is identical to the cutoffs in Figure 27 since the maximum order along the x-axis remains $m = 2$. In each curve there is also a peak that forms around $q = 1$ before flattening out. At this point the central DC component is no longer felt in the signal, which drastically reduces the cross-talk since this central component carries such a strong amplitude, hence any overlap results in large cross-talk. The absence of the DC term therefore instantly reduces the cross-talk.

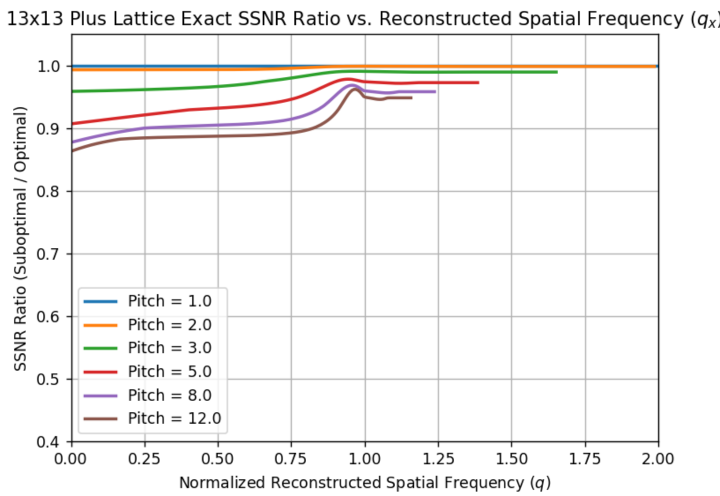

Figure 28: This figure shows how the SSNR ratio varies over the normalised spatial frequencies q for a range of different effective pitch values for the plus shaped lattice. These pitch values are the normalised $p = \frac{1}{\rho}$. Note that this plot is taken along the q_x axis.

6 Unit Cell Lattice SSNR

6.1 Grid of Unit Cells

Another approach to SIM is to take a grid of unit cells, where in the centre of each unit cell there is a focussed spot. The grid is then configured by placing multiple unit cells next to each other. This is visually shown in Figure 29. The distance between spots inside the unit cells is given by the pitch, p and the width of each spot has magnitude of approximately $\frac{\lambda}{NA}$, ensuring that $p \gg \frac{\lambda}{NA}$ a light lattice can be formed. By systematically altering the distance p, the density of the delta peaks can be adjusted within the fixed circular boundary of the back focal plane. In k-space the back focal plane has maximum radius $k_{max} = \frac{NA}{\lambda}$, meaning the BFP has area $A_{\text{BFP}} = \pi \left(\frac{NA}{\lambda}\right)^2$.

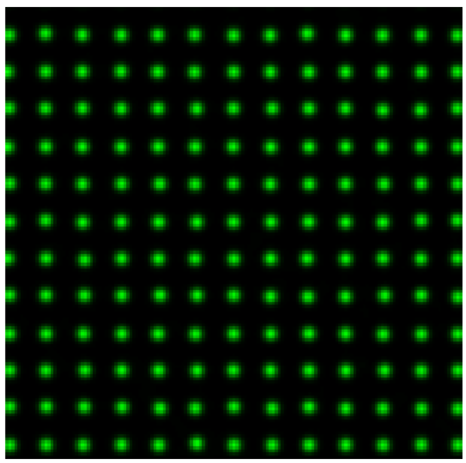

Figure 29: This figure shows the configuration of a unit cell of point sources, arranged to create a lattice that can be used for SIM. Image sourced from [35].

By adjusting the pitch p the amount of point sources that fit into the BFP can be altered. In k-space the changes to p have an inverse effect on the lattice, meaning that in k-space the spots have a periodicity of $\frac{1}{p}$. For example, increasing p would decrease the distance between the points in k-space. The first interesting cases to be considered are $p \to \infty$ and $p \to 0$. For $p \to \infty$, in k-space the points are packed infinitely close together, resulting in the effective light lattice being one massive light spot in the BFP. On the other end with $p \to 0$ the resulting lattice in k-space is that there is only a singular spot on the BFP since the points are spread very far apart and pushed outside the bounds of the BFP.

6.2 Effects of Varying the Pitch

It is of interest to assess in between the extremes values of pitch, $1 \leq p \leq 20$. First, consider large values of p. This gives that in the BFP the spatial frequencies of the point sources are closely packed and located close to each other. As a result, the vector difference between Fourier orders will be smaller and hence evaluating the OTF at these small distances will still be significant. There will be more cross-talk, which will be reflected in the larger off-diagonal elements of the matrix M. It is therefore expected that the simplification $M = I$ is less valid for larger p. In contrast, for small values of p, the opposite effect is observed, when the points in k-space are located further apart. Hence, the vector differences between Fourier orders is larger, and evaluating the OTF at these larger distances falls off quickly to 0. Consequently, the central diagonal dominates and the assumption that $M = I$ would likely be a reasonable approximation. This reiterates the point that although increasing the number of sources may lead to more Fourier orders and therefore more observable spatial frequencies, it comes with the significant draw back that there is much more cross-talk between orders that need to be disentangled.

To visualise this, the SSNR ratio at the spatial frequency $q = 0$ is plotted as a function of the pitch. In the continuous limit of an infinitely dense grid, the Fourier order amplitudes will converge to the analytical incoherent OTF of the circular aperture. However, the unit cell lattice will generally have a finite number of point sources. As the pitch varies, new light spots cross into the boundary of the pupil aperture, which causes a sudden change in the interference energy and cross-talk matrix dimensions. To accurately capture these boundary effects, the Fourier order amplitudes must be calculated by computing the autocorrelation of the unit cell grid. Doing so results in the same lattice structure but with newly computed amplitudes per peak. This is plotted in Figure 30, which reveals a bumpy curve representing the sudden addition of beams as the lattice becomes more dense.

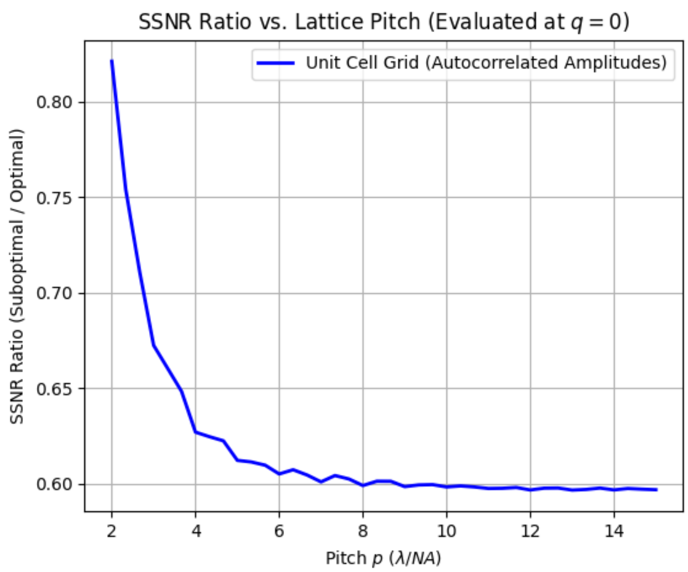

Figure 30: This figure shows a plot of the SSNR ratio vs the pitch p in orders of $\frac{\lambda}{NA}$. It displays that for higher pitch with a more dense arrangement in k-space, the ratio decreases and the performance of the suboptimal assumption $M = I$ is worse. The amplitudes of the Fourier orders are numerically computed, which introduces a ragged shape to the curve.

In Figure 30 it can be seen that when increasing the pitch, and thus the number of unit cells and light spots within the lattice, results in the simplification of $M = I$ becoming worse. In view of computational cost, using M has drawbacks because the denser the lattice, the more points there are that are needed to be considered and therefore computation requires more effort. As an example, consider the case for $p = 15$ units of $\frac{\lambda}{NA}$. In k-space the normalised distance between points is $\frac{1}{15}$, and the aperture radius $k_{max} = 1$, this means that one can fit 29 points along the central horizontal and 29 points along the central vertical. Approximating the circular aperture to a square, this would mean a total of $29 \times 29 = 841$ different light point on the lattice, resulting in a 841×841 matrix being required to properly represent the noise matrix. This takes a significant amount of time to calculate.

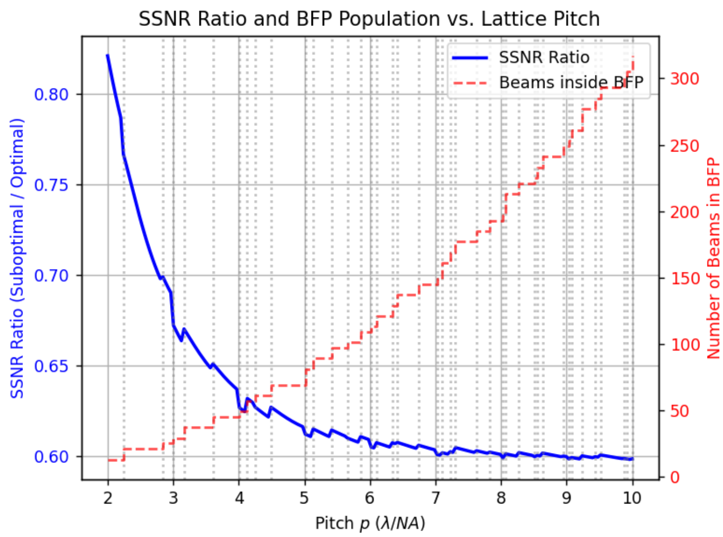

Figure 31: This figure shows a plot of the SSNR ratio and back focal plane (BFP) population vs the pitch p in orders of $\frac{\lambda}{NA}$. This plot has a fine x-axis scale and included more data points. In this plot it can be observed that the jumps in the graph occur exactly at discontinuities in the BFP population.

In order to evaluate the jumps in Figure 30 a new plot has been made that includes more data points over a shorter range, cutting off the higher pitches to ensure the code would not take too long to run. In addition, a plot of the number of beams passing through the circular aperture is plotted vs the pitch, where at each discontinuity jump a grey dotted line is plotted. This is shown in Figure 31. It can be observed that every jump in the SSNR ratio coincides perfectly with a step in the number of beams in the BFP.

There are a couple key observations to be made from Figure 31. Firstly, in between jumps in the number of beams in the BFP, the SSNR ratio is generally downward sloping, as is expected since as the pitch increases, the cross-talk increases. This makes the suboptimal filter a worse approximation. Secondly, it can be observed that these jumps both lie on small peaks and in dips. This is likely due to two mechanisms that work against each other. The increase in SSNR ratio can be due to the new energy added to the system when a new beam is added to the lattice, hence increasing the DC component a_0 and when normalising the amplitudes this therefore reduces the off centre amplitudes, making the noise matrix more diagonally dominant, supporting the $M = I$ simplification. The decrease in SSNR at the jumps is due to having more beams, which leads to more interfering pairs, and therefore more cross-talk.

6.2.1 Evaluation of the Unit Cell SSNR Ratio at Different Spatial Frequencies

The plots in Figures 30 and 31 have been taken specifically at the spatial frequency $q = 0$, however each spatial frequency will have its own SSNR. It is expected that the SSNR will differ at different spatial frequencies, which is calculated and plotted for a range of pitch values in Figure 32. Figure 32 shows the generally expected result that with increasing pitch, the SSNR ratio is less. Another generally expected trend is that the ratio increases for increasing spatial frequencies, which is in line with the results found for the previously derived lattice structures. The most noticeable and interesting difference here is the bumpy pattern observed. At intervals of $q = \frac{1}{p}$, dips in the ratio are observed. Since the Fourier orders are located at intervals of $\frac{1}{p}$, these spatial frequency locations are precisely the points where there is maximum cross-talk, reducing the accuracy of the suboptimal M.

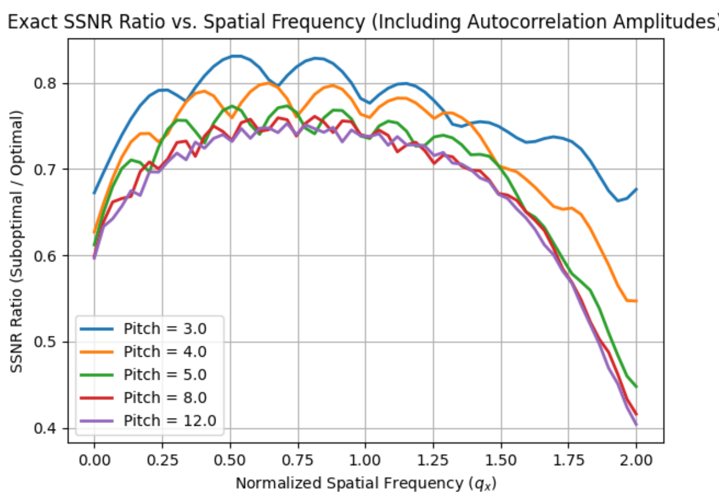

Figure 32: Exact numerical SSNR ratio plotted against normalized spatial frequency, q_x, for a grid of unit cells at various pitch values (p). The plot uses numerically derived autocorrelation amplitudes and is plotted for spatial frequencies along the x-axis. The distinct dips in the SSNR ratio occur at spatial frequencies corresponding to $q_x = n/p$ (for n integers).

As the position of these bumps change with azimuthal angle of $\vec{q}$, it is convenient to loop over the azimuthal angles and compute an average SSNR. This can help smooth out the bumps caused by only looking at a single azimuthal direction, as in Figure 32. This results in Figure 33. In addition, the SSNR ratio improves with increasing spatial frequencies, before starting to dip off. The initial rapidly increasing SSNR is expected as with the previously derived lattice shapes since for lower spatial frequencies there is much more cross-talk between orders since then the OTFs will overlap significantly, reducing the accuracy of the suboptimal filter. The unique element of this figure is that this curve peaks and then drops off for higher spatial frequencies. The likely reason for this is that at these larger shifted spatial frequencies the OTF is much weaker, therefore with the suboptimal filter more signal is lost compared to an optimal filter that would retain more of the already minimal signal. With smaller signal strengths, this increased precision will have larger effects than for stronger signals.

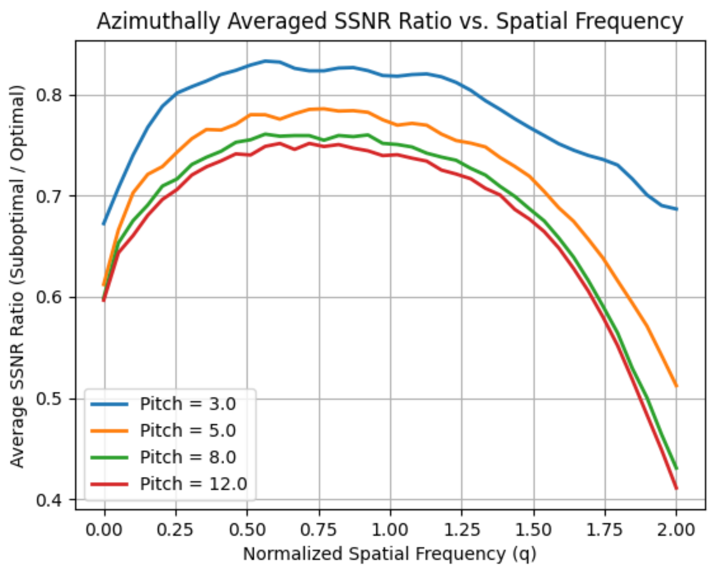

Figure 33: This figure shows the azimuthally averaged SSNR ratio plotted against normalized spatial frequency q for various lattice pitch values. By averaging over all azimuthal angles, the bumps observed in Figure 32 are smoothed out.

6.2.2 SSNR Ratio Convergence and Maxima

In Figures 30 and 31 it can be observed that when increasing the pitch, the SSNR ratio flattens out and stops decreasing as rapidly as it does for smaller pitch values. These curves give the impression that the SSNR ratio converges to a specific value when increasing the pitch beyond a certain point. Taking reference to Figure 30, the SSNR ratio at a $q = 0$ seems to converge to just below 0.6. It can be interesting to assess why this appears to converge and what values of convergence is expected.

As explained in earlier sections, increasing the pitch decreases the distance between orders in k-space. Therefore as the pitch is increased, the resulting illumination pattern observed in k-space becomes more blurred into a singular continuous light spot. At smaller pitch values the orders are more distinct and therefore the SSNR ratio is directly impacted by changes, however at the larger pitches the BFP is already densely populated such that more orders have a more limited impact on the SSNR ratio, which explains why convergence can be expected for increasing pitch.

To get an idea of the convergence value the noise matrix can be assessed. This noise matrix is composed of a product of amplitude and OTF evaluation at a spatial frequency. For large pitch values, the orders in k-space are close together and therefore the vector differences between orders is small and when evaluating the OTF at these vector differences the magnitude is still large. This invalidates the $M = I$ assumption that assumes that the OTF falls off quickly and the OTF evaluation at the vector differences can be assumed to be negligible. The convergence value can be seen as a floor for how poorly the suboptimal filter will perform.

A potential approach for studying the convergence value is taking $p \to \infty$. The light spots in real space are then spaced infinitely far apart, hence it could be seen as a single light spot being used to illuminate the sample. Analogously, in k-space this results in a continuously filled circle illuminating the back focal plane. This concept is identical to image scanning microscopy (ISM) [36]. It would therefore be a logical step to compare the SSNR ratios for ISM and the convergence values for lattice SIM.

To accurately model ISM the discrete model must be mapped to a continuous limit. For this it is important that the scaling of the physical support is preserved. In the discrete model, the BFP has a normalised radius of 1.0, however the illumination amplitudes $\hat{a}_m$ are derived through autocorrelations of the grid in k-space, which has a physical support up to a maximum radius of 2.0. If one therefore assumes that the continuous illumination amplitude is equal to the OTF then an artificial scaling error is introduced. The adjusted scaling to address this is shown in Equation 91.

$$\hat{a}_{cont}(\vec{k}) = \hat{g}\left(\frac{\vec{k}}{2}\right) \tag{91}$$

Using this properly scaled amplitude function, the discrete vectors and noise matrix can be converted into their continuous integral forms, shown in Equations 92 and 93.

$$y(\vec{k}) = \hat{a}_{cont}(\vec{k})\hat{g}(\vec{q} - \vec{k}) \tag{92}$$

$$M(\vec{k}, \vec{k}') = \hat{a}_{cont}(\vec{k} - \vec{k}')\hat{g}(\vec{k}' - \vec{k}) \tag{93}$$

The SSNR ratio is defined analogously to the discrete case, shown in Equation 94.

$$\frac{\text{Suboptimal SSNR}}{\text{Optimal SSNR}} = \frac{|\langle y|y\rangle|^2}{\langle y|M^{-1}|y\rangle\langle y|M|y\rangle} \tag{94}$$

In the continuous limit, the operator M acts on a vector $x(\vec{k})$ in the form of a Fredholm integral equation of the first kind [37], shown in Equation 95.

$$y(\vec{k}) = (Mx)(\vec{k}) = \int M(\vec{k}, \vec{k}')x(\vec{k}')d\vec{k}' \tag{95}$$

Solving this Fredholm integral directly for the inverse term $\langle y|M^{-1}|y\rangle$ is an ill-posed numerical deconvolution problem, which inaccurately represents the SSNR ratio at higher spatial frequencies, meaning solving this problem numerically in this matter will not work. Instead, by inspecting Equation 95, observe that it describes a 2D spatial convolution. By defining a combined kernel $K(\vec{k}) = \hat{a}_{cont}(\vec{k})\hat{g}(\vec{k})$, the integral simplifies to the convolution expression in Equation 96.

$$(My)(\vec{k}) = (K * y)(\vec{k}) \tag{96}$$

This convolution in k-space is equivalent to multiplication in real space, which greatly simplifies the matrix inversion step. The inverse Fourier transforms of the kernel, $K(\vec{k})$, and the continuous function, $y(\vec{k})$, are shown in Equations 97 and 98.

$$P(\vec{r}) = \mathcal{F}^{-1}\{K(\vec{k})\} \tag{97}$$

$$y(\vec{r}) = \mathcal{F}^{-1}\{y(\vec{k})\} \tag{98}$$

Using Parseval's theorem [27], the inner products can be easily computed as real spatial integral as shown in Equations 99 to 101.

$$\langle y|y\rangle = \int |y(\vec{r})|^2 d\vec{r} \tag{99}$$

$$\langle y|M|y\rangle = \int |y(\vec{r})|^2 P(\vec{r})d\vec{r} \tag{100}$$

$$\langle y|M^{-1}|y\rangle = \int \frac{|y(\vec{r})|^2}{P(\vec{r})}d\vec{r} \tag{101}$$

To model the ISM limit SSNR Equations 99, 100 and 101 are modelled in python to produce Figure 34. This figure gives some notable results. Firstly, at $q = 0$, the ISM limit SSNR ratio gives a value of just below 0.6, which is exactly the same as predicted in Figure 33 for $p = 12$. The second notable result is that the curve otherwise looks very similar to the family of curves in Figure 33, just shifted even further down, having a worse suboptimal filter performance. This result is inline with expectations, as the trend shows that for increasing pitch, the SSNR ratio tends to gradually becomes smaller. To illustrate how the curves continue to decrease towards the found ISM limit curve, a plot including a larger pitch of $p = 20$ is plotted, giving the result in Figure 34.

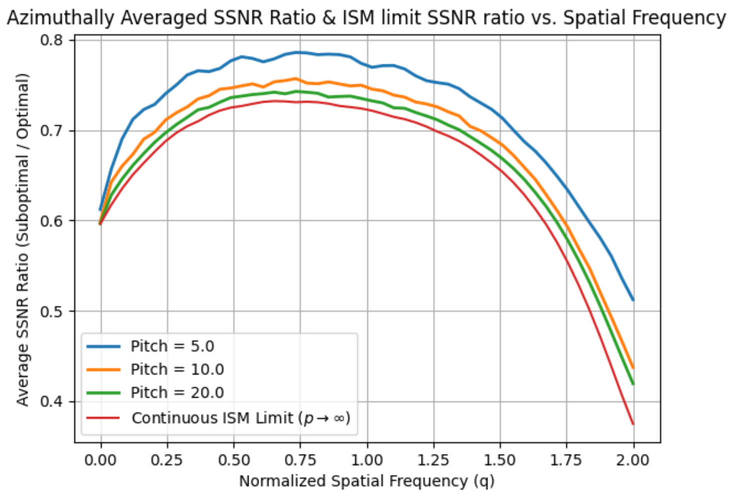

Figure 34: This figure shows the SSNR ratio vs normalised spatial frequencies for the limit case of lattice SIM taking the pitch to infinity. It can be observed that the curves indeed appear to converge to the found ISM limit case.

Moreover, an interesting aspect of Figures 33 and 34 is that each curve appears to have a maxima around the spatial frequency $q = 0.6$. A potential physical explanation for the observed maxima can be found by determining the causes for a rise and decline in the SSNR ratio. The rise in SSNR ratio is due to the mathematical penalty in omitting the cross-talk when using $M = I$, which is most profoundly felt at lower spatial frequencies. The fall in SSNR ratio at the higher SSNR ratios come from the absolute signal loss at the higher spatial frequencies, where every extra detail obtained though the optimal filter becomes more significant at higher spatial frequencies. The maxima would occur precisely in the intersection where the mathematical cross-talk penalty at lower spatial frequencies and the absolute signal loss effects at higher spatial frequencies intersect.

This seeming maxima occurring around $q = 0.6$ is a familiar number in optics, since at first sight it appears to be no coincidence and potentially related to the Abbe diffraction limit $d = 0.61\frac{\lambda}{NA}$, with d the minimum resolvable distance between two points. Unfortunately, these are not directly related. This is because the Abbe diffraction concerns the real space, whilst the plot in Figure 33 works with spatial frequencies in k-space. Since k-space is a reciprocal space to real space, a more related Abbe diffraction expression would be $f_{Abbe} = \frac{NA}{0.61\lambda}$. The x-axis is normalised to $f_{cut} = \frac{2NA}{\lambda}$, hence $\rho_{Abbe} = \frac{f_{Abbe}}{f_{cut}} \approx 0.82$. Should this diffraction limit be responsible, then the artifact is expected around 0.82.

7 Conclusion

Structured illumination microscopy (SIM) is a super resolution technique that bypasses the diffraction limit by downshifting higher spatial frequency information within the optical transfer function (OTF) support. By taking multiple images at different illumination phase shifts, the SIM reconstruction algorithm can recombine these images to form a single high-detail image. Within the reconstruction algorithm, it is a common convention to neglect the cross-talk between Fourier orders, simplifying the so-called noise matrix, M, to the identity matrix, $M = I$. This thesis set out to assess the validity of the assumption $M = I$ and to evaluate the necessity for working directly with the full noise matrix instead.

The noise matrix is made up of a central diagonal, the DC terms of the light signal, and the off-diagonals representing the cross-talk between light sources. The simplification assumes that there is minimal interference between light beams in the lattice and thus that the cross-talk terms are sufficiently small to be neglected during reconstruction. The first notable point regards analytical bounds. A middle ground approach could potentially be viable as an alternative to working with M during reconstruction if suitable bounds on the SSNR ratio could be computed such that these bounds can be used in a way that does not require hard calculations. For example, perhaps with a bound value, b, the simplification could yield that $x = by$ rather needing to use $x = M^{-1}y$. The bounds considered in the project were the Kantorovich bound and the Bounding Theorem for Toeplitz matrices. Both of these bounds are unsuitable for application within the reconstruction algorithm, since these bounds come with significant drawbacks. Firstly, the Kantorovich bound is extremely loose, resulting in an overestimate for the penalty incurred. Consequently, this bound became very unrepresentative of the true SSNR ratio. Another drawback to this method is that it requires the knowledge of the minimum and maximum eigenvalues for the noise matrix M. In many cases, the computation of these eigenvalues is challenging and must be done numerically anyway. Since numerical calculation is required, it is far more constructive to directly compute the SSNR ratio numerically as well, rather than working with bounds that are very loose. The approach to remove the requirement of numerical determination of the eigenvalues is to use the Bounding Theorem. This theorem, however, has the drawback that this gives an inequality on top of the Kantorovich inequality. As a result, this bound becomes extremely loose and often highly misrepresentative. A further drawback of the Bounding Theorem is the requirement that M is Toeplitz. Unfortunately, M quickly loses its Toeplitz structure, requiring an additional mathematical step to restore the matrix to a BTTB structure. Regardless, by this point the bounds have become so loose that it provides no additional insights on the true SSNR ratio, leading to the conclusion that evaluating the SSNR ratio is best done numerically.

Moreover, by computing the SSNR ratios numerically the necessity for the more rigorous approach using M directly in the reconstruction can be concluded. For the 2D conventional SIM the SSNR ratio remains high, with values above 0.9 for spatial frequencies $0.0 \leq q \leq 0.8$. Even beyond this range the SSNR ratio remains large, never dropping below 0.7. For this lattice configuration working with the assumption that $M = I$ is perfectly valid. Similarly, when adding the third beam to achieve 3D conventional SIM, the SSNR ratio almost remains above 0.9 for spatial frequencies $0.0 \leq q \leq 0.8$. The assumption for this case does however begin to collapse for larger pitch values at larger spatial frequencies. The SSNR ratio dips down to below 0.65, suggesting room for a 35% improvement that can be achieved. This lattice configuration would therefore benefit from using the precise M matrix since there is noticeable cross-talk for the dense lattice arrangements. In contrast, the plus lattice retains SSNR ratios above 0.85 for all spatial frequencies and pitches. This means that the simplification step $M = I$ is a more valid approximation for this lattice configuration. This validity stems from the centrally dominated noise matrix for the lattice since the DC component is relatively large. Since SIM applications tend to predominantly use these three lattice types, it is generally acceptable to use the simplification assumption.

For the unit cell lattice we draw a different conclusion however. For almost all pitch values, the SSNR ratio at the zero spatial frequency is already low, with values around 0.6 to 0.7. This low ratio at $q = 0$ is due to the large amount of cross-talk directly at the centre due to the overlapping OTFs of surrounding unit cells. Since such a large improvement can be achieved directly within the non-super resolved range already, it is clear that for this lattice arrangement it is vital to use the correct noise matrix within the reconstruction algorithm. This conclusion is further strengthened when assessing the super resolved regime for spatial frequencies above 1.0. Since SIM attempts to resolve those frequencies, it is essential that a clear image is made at those spatial frequencies. Since the SSNR ratio drops off rapidly for increasing spatial frequencies and for each pitch, it becomes apparent that the penalty incurred is significant. The potential improvement reaches upwards of 50% for the extreme spatial frequencies near cutoff.

Further, taking the pitch to infinity in the unit cell configuration results in orders blurring together into one continuous pattern full of overlap. This limit case has been found to be an alternative super resolution technique called ISM. The unit cell SNNR ratios converges to this limiting case SSNR ratio. In this limit, using the suboptimal filter reduces the quality significantly, with the SSNR ratio remaining below 0.75 for all spatial frequencies and dropping below 0.40 at near cutoff. This means a minimum improvement of 25% can be achieved and a maximum of 60% can be achieved, which is a substantial improvement.

To conclude, the simplification $M = I$ almost always results in a worse performing system, however for specific use cases of lattice type and pitch the penalty is less drastic. For these instances, working with this approximation can be beneficial to reduce the computation steps and time. For the other use cases with large pitch or with a large amount of sources in the lattice, the cross-talk becomes highly significant and neglecting the interactions between orders is heavily penalised. Here a large amount of improvement can be brought to the image quality, making it advisable to consider rigorously using the noise matrix during reconstruction. A suitable follow up question could be to analyse the time constraint problem related to fully computing the noise matrix in the algorithm. Given the necessity for quick imaging using SIM, the additional computation time must be taken into consideration to assess the viability of this additional step. It can therefore be helpful to consider different approaches to compare using $M = I$, using M directly or using a some middle ground approach that requires less computation, but that still considers some cross-talk terms to improve the SSNR ratio.

8 References

[1] Nature Photonics. Beyond the diffraction limit. *Nat. Photonics*, 3(7):361, 2009.

[2] S. T. Ross and M. W. Davidson. Introduction to fluorescence microscopy. In *Methods in Cell Biology*, volume 123, pages 1–18. Academic Press, 2014.

[3] A. P. Konijnenberg, A. J. L. Adam, and H. P. Urbach. *BSc Optics*. TU Delft OPEN, 2nd edition, 2024.

[4] L. Schermelleh, R. Heintzmann, and H. Leonhardt. A guide to super-resolution fluorescence microscopy. *J. Cell Biol.*, 190(2):165–175, 2010.

[5] K. Samanta and J. Joseph. An overview of structured illumination microscopy: Recent advances and perspectives. *J. Opt.*, 23(12):123002, 2021.

[6] J. Alvelid and I. Testa. Fluorescence microscopy at the molecular scale. *Curr. Opin. Biomed. Eng.*, 12:34–42, 2019.

[7] W. Lukosz. Optical systems with resolving powers exceeding the classical limit. *J. Opt. Soc. Am.*, 56(11):1463–1471, 1966.

[8] R. Heintzmann, T. M. Jovin, and C. Cremer. Saturated patterned excitation microscopy–a concept for optical resolution improvement. *J. Opt. Soc. Am. A*, 19(8):1599–1609, 2002.

[9] V. Saveljev, S.-K. Kim, and J. Kim. Moiré effect in displays: a tutorial. *Opt. Eng.*, 57(3):030803, 2018.

[10] Y. Ma, K. Wen, M. Liu, J. Zheng, K. Chu, Z. J. Smith, L. Liu, and P. Gao. Recent advances in structured illumination microscopy. *J. Phys. Photonics*, 3(2):024009, 2021.

[11] J. Kalkman and K. Grussmayer. *Applied Optical Imaging*. Delft University of Technology, 2023. Accessed: 2026-06-15.

[12] J. Valli, A. Garcia-Burgos, L. M. Rooney, B. Vale de Melo E Oliveira, R. R. Duncan, and C. Rickman. Seeing beyond the limit: A guide to choosing the right super-resolution microscopy technique. *J. Biol. Chem.*, 297(1):100791, 2021.

[13] H. Wu, Y. Li, Y. Sun, L. Yin, W. Sun, and Z. Ye. Single-frame structured illumination microscopy for fast live-cell imaging. *APL Photonics*, 9(3):036102, 2024.

[14] N. de Jonge, D. B. Peckys, G. J. Kremers, and D. W. Piston. Electron microscopy of whole cells in liquid with nanometer resolution. *Proc. Natl. Acad. Sci. U.S.A.*, 106(7):2159–2164, 2009.

[15] E. F. Garman and M. Weik. Radiation damage to biological samples: still a pertinent issue. *J. Synchrotron Radiat.*, 28(6):1621–1623, 2021.

[16] K. W. Kim. Some living eukaryotes during and after scanning electron microscopy. *Appl. Microsc.*, 51(1):16, 2021.

[17] F. Ströhl and C. F. Kaminski. Frontiers in structured illumination microscopy. *Optica*, 3(6):667–677, 2016.

[18] CrestOptics. Spinning disk structured illumination microscopy (SD-SIM) as a success key for fast live-cell imaging. Application note, CrestOptics, 2022. Accessed: 2026-06-15.

[19] C. S. Smith, J. A. Slotman, L. Schermelleh, J. Gražvydas, B. Rieger, and S. Stallinga. Structured illumination microscopy with noise-controlled image reconstructions. *Nat. Methods*, 18(7):821–828, 2021.

[20] R. C. Gonzalez, R. E. Woods, and B. R. Masters. Digital image processing, third edition. *J. Biomed. Opt.*, 14(2):029901, 2009.

[21] S. Stallinga. Imaging systems: Structured illumination microscopy. Lecture slides, Department of Imaging Physics, TU Delft, n.d.

[22] M. G. L. Gustafsson. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.*, 198(2):82–87, 2000.

[23] Y. Liu. *Optical super-resolution techniques for 3D profilometry*. PhD thesis, Universitat Politècnica de Catalunya, 2019. UPCommons institutional repository.

[24] B. Morrison and Y. Xu. Super-resolution optical microscopy. Ultrafast and Microspectroscopy Laboratories, The University of Melbourne, n.d. Accessed: 2026-06-17.

[25] V. Brudanin, B. Rieger, and S. Stallinga. Comparison of 3d structured illumination microscopy configurations in terms of spectral signal to noise ratio. *Opt. Express*, 33(8):12803–12821, 2025.

[26] V. Brudanin. SIM: Repository for the code on structured illumination microscopy. `https://github.com/valbrud/SIM/tree/master`, 2025. GitHub repository.

[27] J. W. Goodman. *Introduction to Fourier Optics*. W. H. Freeman and Company, 4th edition, 2017.

[28] M. Gu and D. Bird. Three-dimensional optical-transfer-function analysis of fiber-optical two-photon fluorescence microscopy. *J. Opt. Soc. Am. A*, 20(5):941–947, 2003.

[29] M. Gu. *Advanced Optical Imaging Theory*. Springer Series in Optical Sciences. Springer Berlin Heidelberg, 2000.

[30] U. Bal. Dual tree complex wavelet transform based denoising of optical microscopy images. *Biomed. Opt. Express*, 3(12):3231–3239, 2012.

[31] S. W. Hasinoff. Photon, Poisson noise. In *Computer Vision: A Reference Guide*, pages 608–610. Springer US, 2014.

[32] D. Gómez. A more direct proof of gerschgorin's theorem. *Matemáticas: Enseñanza Universitaria*, 14(2):119–122, 2006.

[33] M. Newman. Kantorovich's inequality. *J. Res. Natl. Bur. Stand. B*, 64B(1):33–34, 1960.

[34] R. M. Gray. Toeplitz and circulant matrices: A review. *Found. Trends Commun. Inf. Theory*, 2(3):155–239, 2006.

[35] N. Chakrova. *Versatile Structured Illumination Microscopy*. PhD thesis, Delft University of Technology, 2016. TU Delft Institutional Repository.

[36] C. B. Müller and J. Enderlein. Image scanning microscopy. *Phys. Rev. Lett.*, 104(19):198101, 2010.

[37] G. B. Arfken, H. J. Weber, and F. E. Harris. *Mathematical Methods for Physicists: A Comprehensive Guide*, chapter 21, pages 1011–1052. Academic Press, 7th edition, 2013.

Appendix

Illumination Pattern Magnitudes

The following tables giving the positions and magnitudes of the illumination pattern for various lattice shapes have been sourced from [25] or manually derived. In these magnitudes, the amplitude parameter r is defined as the ratio of the amplitude of the obliquely incident plane wave component and the x or y component of the normally incident plane wave component.

Table 1: Fourier structure for the illumination pattern resulting for 2D conventional SIM with two point sources.

Position	**Magnitude**
$(0,0)$	$1+r^2$
$(\pm 1,0)$	r

Table 2: Fourier structure for the illumination pattern resulting for 3D conventional SIM with three point sources.

Position	**Magnitude**
$(0,0)$	$1+2r^2$
$(\pm 1,0)$	$2r$
$(\pm 2,0)$	r^2

Table 3: Fourier structure for the illumination pattern resulting from the plus shaped light lattice with linearly polarised plane waves. Table sourced from [25].

Position	**Magnitude**	**Position**	**Magnitude**
$(0,0,0)$	$2+4r^2$	$\pm(0,1,1)$	$-r$
$(\pm 2,0,0)$	$-r^2$	$\pm(0,1,-1)$	r
$(0,\pm 2,0)$	$-r^2$	$\pm(1,0,1)$	r
		$\pm(1,0,-1)$	$-r$

Table 4: Fourier structure for the illumination pattern resulting from the plus shaped light lattice with circularly polarised plane waves. Table sourced from [25].

Position	**Magnitude**	**Position**	**Magnitude**
$(0,0,0)$	$2+8r^2$	$\pm(1,0,1)$	$-(1+\cos\theta)r$
$(\pm 2,0,0)$	$-2r^2\cos^2\theta$	$\pm(-1,0,1)$	$(1+\cos\theta)r$
$(0,\pm 2,0)$	$-2r^2\cos^2\theta$	$(0,1,\pm 1)$	$-i(1+\cos\theta)r$
$(\pm 1,\pm 1,0)$	$-2r^2\sin^2\theta$	$(0,-1,\pm 1)$	$i(1+\cos\theta)r$

Table 5: Fourier structure for the illumination pattern resulting from the hexagonally shaped light lattice with linearly polarised plane waves. Table sourced from [25].

Position	**Magnitude**	**Position**	**Magnitude**
$(0,0,0)$	$2+6r^2$	$(1/2, \sqrt{3}/2, \pm 1)$	$\frac{\mp\sqrt{3}-i}{2}r$
$(\pm 2,0,0)$	$-r^2$	$(-1/2, \sqrt{3}/2, \pm 1)$	$\frac{\mp\sqrt{3}+i}{2}r$
$(\pm 1, \pm\sqrt{3}, 0)$	$-r^2$	$(1/2, -\sqrt{3}/2, \pm 1)$	$\frac{\pm\sqrt{3}-i}{2}r$
$(0, \pm\sqrt{3}, 0)$	$-r^2$	$(-1/2, -\sqrt{3}/2, \pm 1)$	$\frac{\pm\sqrt{3}+i}{2}r$
$(\pm 3/2, \pm\sqrt{3}/2, 0)$	$-r^2$	$(1,0,\pm 1)$	$-ir$
$(\pm 1, 0, 0)$	r^2	$(-1,0,\pm 1)$	ir
$(\pm 1/2, \pm\sqrt{3}/2, 0)$	r^2		

Full Noise Matrix Representation

For the plus shape lattice with linearly s-polarised plane waves with amplitudes and spatial frequencies given by Table 3 the matrix M is given by Equation 102, where entries $M_{i,j} = c(\vec{q}_i - \vec{q}_j) = \hat{a}_{\vec{q}_i - \vec{q}_j}\hat{g}(\vec{q}_j - \vec{q}_i)$.

$$M_{plus} = \begin{bmatrix} c_{0,0,0} & c_{-2,0,0} & c_{2,0,0} & c_{0,-2,0} & c_{0,2,0} & c_{0,-1,-1} & c_{0,1,1} & c_{0,-1,1} & c_{0,1,-1} & c_{-1,0,-1} & c_{1,0,1} & c_{-1,0,1} & c_{1,0,-1} \\ c_{2,0,0} & c_{0,0,0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{1,0,-1} & 0 & c_{1,0,1} & 0 \\ c_{-2,0,0} & 0 & c_{0,0,0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{-1,0,1} & 0 & c_{-1,0,-1} \\ c_{0,2,0} & 0 & 0 & c_{0,0,0} & 0 & c_{0,1,-1} & 0 & c_{0,1,1} & 0 & 0 & 0 & 0 & 0 \\ c_{0,-2,0} & 0 & 0 & 0 & c_{0,0,0} & 0 & c_{0,-1,1} & 0 & c_{0,-1,-1} & 0 & 0 & 0 & 0 \\ c_{0,1,1} & 0 & 0 & c_{0,-1,1} & 0 & c_{0,0,0} & 0 & 0 & c_{0,2,0} & 0 & 0 & 0 & 0 \\ c_{0,-1,-1} & 0 & 0 & 0 & c_{0,1,-1} & 0 & c_{0,0,0} & c_{0,-2,0} & 0 & 0 & 0 & 0 & 0 \\ c_{0,1,-1} & 0 & 0 & c_{0,-1,-1} & 0 & 0 & c_{0,2,0} & c_{0,0,0} & 0 & 0 & 0 & 0 & 0 \\ c_{0,-1,1} & 0 & 0 & 0 & c_{0,1,1} & c_{0,-2,0} & 0 & 0 & c_{0,0,0} & 0 & 0 & 0 & 0 \\ c_{1,0,1} & c_{-1,0,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{0,0,0} & 0 & 0 & c_{2,0,0} \\ c_{-1,0,-1} & 0 & c_{1,0,-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{0,0,0} & c_{-2,0,0} & 0 \\ c_{1,0,-1} & c_{-1,0,-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{2,0,0} & c_{0,0,0} & 0 \\ c_{-1,0,1} & 0 & c_{1,0,1} & 0 & 0 & 0 & 0 & 0 & 0 & c_{-2,0,0} & 0 & 0 & c_{0,0,0} \end{bmatrix} \tag{102}$$

Toeplitz Blocks

For the BTTB plus shaped light lattice more of the Toeplitz blocks are given by the following matrices.

$$T_1 = \begin{pmatrix} a_{0,1}g(0,k) & a_{1,1}g(k,k) & a_{2,1}g(2k,k) & a_{3,1}g(3k,k) & a_{4,1}g(4k,k) \\ a_{-1,1}g(-k,k) & a_{0,1}g(0,k) & a_{1,1}g(k,k) & a_{2,1}g(2k,k) & a_{3,1}g(3k,k) \\ a_{-2,1}g(-2k,k) & a_{-1,1}g(-k,k) & a_{0,1}g(0,k) & a_{1,1}g(k,k) & a_{2,1}g(2k,k) \\ a_{-3,1}g(-3k,k) & a_{-2,1}g(-2k,k) & a_{-1,1}g(-k,k) & a_{0,1}g(0,k) & a_{1,1}g(k,k) \\ a_{-4,1}g(-4k,k) & a_{-3,1}g(-3k,k) & a_{-2,1}g(-2k,k) & a_{-1,1}g(-k,k) & a_{0,1}g(0,k) \end{pmatrix} \tag{103}$$

$$T_2 = \begin{pmatrix} a_{0,2}g(0,2k) & a_{1,2}g(k,2k) & a_{2,2}g(2k,2k) & a_{3,2}g(3k,2k) & a_{4,2}g(4k,2k) \\ a_{-1,2}g(-k,2k) & a_{0,2}g(0,2k) & a_{1,2}g(k,2k) & a_{2,2}g(2k,2k) & a_{3,2}g(3k,2k) \\ a_{-2,2}g(-2k,2k) & a_{-1,2}g(-k,2k) & a_{0,2}g(0,2k) & a_{1,2}g(k,2k) & a_{2,2}g(2k,2k) \\ a_{-3,2}g(-3k,2k) & a_{-2,2}g(-2k,2k) & a_{-1,2}g(-k,2k) & a_{0,2}g(0,2k) & a_{1,2}g(k,2k) \\ a_{-4,2}g(-4k,2k) & a_{-3,2}g(-3k,2k) & a_{-2,2}g(-2k,2k) & a_{-1,2}g(-k,2k) & a_{0,2}g(0,2k) \end{pmatrix} \tag{104}$$

$$T_3 = \begin{pmatrix} a_{0,3}g(0,3k) & a_{1,3}g(k,3k) & a_{2,3}g(2k,3k) & a_{3,3}g(3k,3k) & a_{4,3}g(4k,3k) \\ a_{-1,3}g(-k,3k) & a_{0,3}g(0,3k) & a_{1,3}g(k,3k) & a_{2,3}g(2k,3k) & a_{3,3}g(3k,3k) \\ a_{-2,3}g(-2k,3k) & a_{-1,3}g(-k,3k) & a_{0,3}g(0,3k) & a_{1,3}g(k,3k) & a_{2,3}g(2k,3k) \\ a_{-3,3}g(-3k,3k) & a_{-2,3}g(-2k,3k) & a_{-1,3}g(-k,3k) & a_{0,3}g(0,3k) & a_{1,3}g(k,3k) \\ a_{-4,3}g(-4k,3k) & a_{-3,3}g(-3k,3k) & a_{-2,3}g(-2k,3k) & a_{-1,3}g(-k,3k) & a_{0,3}g(0,3k) \end{pmatrix} \tag{105}$$

$$T_4 = \begin{pmatrix} a_{0,4}g(0,4k) & a_{1,4}g(k,4k) & a_{2,4}g(2k,4k) & a_{3,4}g(3k,4k) & a_{4,4}g(4k,4k) \\ a_{-1,4}g(-k,4k) & a_{0,4}g(0,4k) & a_{1,4}g(k,4k) & a_{2,4}g(2k,4k) & a_{3,4}g(3k,4k) \\ a_{-2,4}g(-2k,4k) & a_{-1,4}g(-k,4k) & a_{0,4}g(0,4k) & a_{1,4}g(k,4k) & a_{2,4}g(2k,4k) \\ a_{-3,4}g(-3k,4k) & a_{-2,4}g(-2k,4k) & a_{-1,4}g(-k,4k) & a_{0,4}g(0,4k) & a_{1,4}g(k,4k) \\ a_{-4,4}g(-4k,4k) & a_{-3,4}g(-3k,4k) & a_{-2,4}g(-2k,4k) & a_{-1,4}g(-k,4k) & a_{0,4}g(0,4k) \end{pmatrix} \tag{106}$$