Investigating the applicability of LCD technology to structured illumination microscopy

Matthew Caldwell, CoMPLEX, UCL

Supervisors:
Dr Guy Moss, CoMPLEX & Department of Pharmacology, UCL
Dr Sally Day, Department of Electronic & Electrical Engineering, UCL

Word count: 16,300

August 31, 2007

Contents

Abbreviations & symbols

CCD	Charge-coupled device (image sensor)
DMD	Digital micromirror device
GFP	Green fluorescent protein
LCD	Liquid crystal display
LCOS	Liquid crystal on silicon
NA	Numerical aperture
OTF	Optical transfer function
PSF	Point spread function
SI	Structured illumination
SICM	Scanning ion conductance microscopy
SLM	Spatial light modulator
SNR	Signal to noise ratio

$\mathcal{F}$	Fourier transform
$\otimes$	Convolution operator

1 Introduction

Optical microscopy has been central to biological investigation for hundreds of years, and its importance only continues to grow. When combined with the wealth of techniques for introducing and controlling fluorescent probes in living systems, the ability to image those systems non-destructively and under physiological conditions allows vital insights into the cellular processes on which life depends.

However, the optical microscope faces a significant resolution limit due to the wave nature of light, and this constrains the uses to which it can be put. Although higher resolution alternative methods exist, they often cannot be applied in living cells, and tend to be cumbersome and expensive. There is therefore considerable interest in extending the resolution of fluorescence microscopy, and several techniques have been developed to this end.

One promising approach is the use of *structured illumination* (SI), in which knowledge of specific patterns imposed on the excitation light allows multiple images to be combined algorithmically to create a higher resolution composite. Several variations on this idea exist, and have been applied to improve both axial and lateral resolution, but current implementations are too slow to capture the dynamic activities of living cells.

This report describes our investigation into the possibility of implementing SI techniques at higher speed and with greater flexibility by using a high-resolution *liquid crystal display* (LCD) component to generate the illumination patterns. The appeal of such an implementation should be obvious, but there are many physical and operational details to consider.

In §2 below, we give a brief overview of the importance of microscopy to modern experimental biology and provide a rationale for this project. §3 and §4 lay out the theoretical bases of SI microscopy, while the technological background to our approach is explored in §5 and §6. These discussions are synthesised in §7 to present an initial design proposal and assess its practicality, with particular reference to the goals suggested by §2. Finally, §8 outlines possible future work in this direction.

2 Biological significance

2.1 Small world

Living systems operate across an enormous scale range, from the molecular level of individual chemical reactions up to the dynamics of entire populations spanning the globe. In attempting to understand and analyse such systems, the biological sciences are faced with the problem of observing and measuring their activities over those staggeringly disparate scales.

Macroscopic phenomena, while subject to any number of other measurement difficulties, at least admit the possibility of direct apprehension by our senses. For things at smaller scales this is not so. We can devise experiments to help elucidate what occurs at these levels, but in the absence of any mechanism of direct observation our understanding must be incomplete at best.

Aside from a few primitive classes of entity at the very limit of what we consider 'alive', the existence of all living things is based around the *cell*. Cells are the functional units of biology, the packets into which life is partitioned; its *atoms*.[1] The vast majority of them—and certainly

[1] Not, of course, literally indivisible, but—as it turned out—neither are atoms themselves.

the internal processes that drive them—are also far too small for human senses to perceive individually. It is only since the advent of the microscope, with it ability to bring the cellular scale into observable range, that our species has made appreciable inroads into understanding how life *works*. In a real sense, biology as it exists today is a direct product of microscopy.

When it comes to making the minuscule visible, the advantages of the optical microscope are many and well-rehearsed, but probably the most important is that it is not innately *destructive*. Most of the other methods we have for probing the world of the cell require the cell in question to be killed and fixed in place. But we recognise intuitively that one of the essential characteristics of living things is their dynamic nature, which such methods sacrifice. Optical microscopy permits cells to be observed as they go about their business in something akin to the conditions under which they exist in real life.

Simply being able to *see* cells in action, while immensely useful, is not enough in itself. Much of what goes on in a cell looks pretty much the same under ordinary illumination. While various gross features can be distinguished, the intricate details get lost in the murk—and cells are fantastically intricate and detailed things. To really see what's happening in them we need to be able to pick out just the individual elements we're interested in at any given time.

There are a number of different techniques for doing this, but what many of them share is that they expose the intricate details by making them conditionally visible via the medium of *fluorescence*. A fluorescent molecule—or fluorophore—absorbs light of one wavelength and then emits light of a shorter wavelength a short time later. As a result, it glows a different colour when excited. Fluorophores can be added to sub-cellular entities from without, for instance by creating antibodies that attach themselves to the items of interest and which also carry the fluorescent probe; or the cell can be coerced into manufacturing certain protein-based fluorophores itself as an adjunct to the processes we wish to observe. Either way, if the cell is excited with the appropriate light and viewed through a filter that blocks the excitation wavelength but not the fluorescent emissions, the items of interest can be seen lit up like a Christmas tree—and the whole viewing process remains in the nice, safe, cell-friendly and directly-observable optical spectrum.

The combination of optical microscopy and fluorescent probes has become one of the key tools in experimental biology, helping to reveal the world of the cell as a place teeming with complex activity. In the process, many new structures and levels of detail have emerged as objects of study. Inevitably, some of these are at or beyond the limits of what optical microscopy can detect.

2.2 Obstacles to exploration

For all its usefulness, fluorescence microscopy is not magic and is still bound by the laws of physics and by the context in which observations are made. These give rise to important limitations on what can be resolved.

Biological specimens are three dimensional. The images obtained through a microscope are two dimensional and directional—we look at a specimen from a particular viewpoint. We therefore distinguish two categories of spatial resolution: *axial*, along the direction of view, and *lateral*, across the plane the resulting image.

Axial resolution is impaired because we are looking *through*—and also illuminating—a *volume* of the sample. While only a portion of that volume will be *in focus*, fluorescent emissions may just as easily emanate from the unfocused regions, and some of these will be imaged by the microscope. If the only fluorescent object in the sample were the one of interest, this would not

be a problem. Alas, it is not generally possible to organise our experiments with that degree of specificity. There are usually at least *some* out of focus emissions contributing to the image, with a consequent reduction in clarity.[2] The matter of separating these extraneous contributions from those in which we are actually interested is the key problem of axial resolution. Ordinary fluorescence microscopy has little to offer in this regard. SI, on the other hand, can provide significant advantages, as will be described in §4.1 [25, 26].

Even were the observed volume to be infinitesimally thin, there would still be a problem of lateral resolution, for reasons discussed in §3.1. The upshot is that for any observational light wavelength—which is to say, for a given colour of fluorescent emissions—there is a corresponding limit to the amount of detail a particular lens is capable of resolving. As the specimen details get smaller and smaller, the ability of the lens to image them declines. Sufficiently small details are indistinguishable from uniformity. This constraint is what is commonly referred to as the diffraction limit, and breaching it is one of the main goals of SI [13, 14, 15, 17].

A final resolution difficulty is that of *time*. An image, in microscopy or otherwise, is not formed instantaneously. Rather, it is an accumulation over time of information transmitted in tiny packets. If the interval is short, then the amount of information is low and can be dominated by stochastic factors such as thermal noise. Over longer exposures or integration times, such noise tends to average out while meaningful content accumulates, improving the *signal to noise ratio* (SNR).

However, in most biological applications, the system being imaged is also *changing* over time. The degree to which this temporal evolution can be resolved depends on the integration time: the image will bundle together all the changes that take place during this period. A longer exposure will provide poorer time resolution, which can be crucial if we are interested in, for example, the order in which a sequence of events occurs—or even just knowing that they *are* separate events.

There is thus a tension between the demands of image quality and those of temporal resolution. SI does not offer any benefits here—quite the reverse. Since it demands the integration of multiple captured images, it can significantly increase the effective exposure time, with a concomitant loss of temporal detail. This is especially the case for existing opto-mechanical implementations, in which the time resolution is limited not by the exposure time for individual frames but by the time taken *in between frames* to physically move the elements generating the illumination pattern. While some trade-off of poorer time resolution for increased spatial resolution is inherent in SI, these implementations squander a great deal of time unnecessarily. One of the key aims of the current project is to see if this time can be reclaimed by using a more efficient means of pattern generation.

2.3 Application examples

To put the foregoing into some kind of context, we now consider the issues of spatial and temporal resolution as they affect some specific experimental situations. The field of biology is so broad, and fluorescence microscopy so ubiquitous, that these examples can barely scratch

[2] We can view this problem of imaging only the features of interest as a generalisation of that addressed by the use of fluorescent probes in the first place. Unfortunately, it is not amenable to the same solution. With fluorescent probes, we can to a large extent restrict our observations to the phenomena of interest. However, those phenomena still occur within a population of cells, manifesting with unknown likelihood over that population and thus over the volume under observation. Just because we can constrain *what* we see, that doesn't mean we can constrain *where* we see it. The latter is part of the experiment *outcome*, not its setup.

the surface of possible applications, but they should serve to illustrate how observational aims might benefit from an improved SI implementation.

The processes that drive a multicellular organism require a great deal of delicate coordination and communication between cells, and this is mediated by a wide range of signalling chemicals. These chemicals must be manufactured, stored until needed, and then released in appropriate amounts at the appropriate times. One common mechanism for storage and release uses constructs called *vesicles*: small membrane-wrapped packets into which the signalling chemicals are placed. Some or all of the vesicle contents can then be ejected into the extracellular matrix by fusion of the vesicle membrane with the cell's own.

Vesicular exocytosis like this manifests in a range of different contexts, but its workings are not well understood. A case in point is the release of *neuropeptides*—larger, comparatively slow-acting signals that may have effects over long distances within the body. These are packed into large vesicles called *dense core granules* [1]. Unlike the smaller neurotransmitters that mediate direct synaptic communication between neurons—which are generally localised in the synapses—neuropeptides may be released from any compartment of the cell. How this release is distributed and coordinated—even whether the behaviour in different compartments is the same—is currently unknown.

To understand this behaviour properly it would be necessary to image activity in different compartments at the same time. Such imaging is beyond the capabilities of most current microscopy techniques; an exception is very high end scanning laser confocal microscopy, but the capital costs of the necessary equipment to perform this are prohibitively high for routine use. Existing SI implementations could be used in combination with a piezoelectric stage and some custom control software to image separate cell compartments, but the time resolution would be too low to make it worthwhile. The shortfall is not huge—dense core granule fusion events take place on scales of tenths of a second—so even a two-to-fourfold framerate increase would make SI a practical contender. Such an increase is not unrealistic with improved pattern generation.

Dense core granules are also examples of cellular structures that are only a short way beyond the lateral resolving power of unaided optical microscopy. Typical sizes for these granules lie in the 100-150nm range—significantly larger, for example, than the vesicles carrying short range neurotransmitters. The resolution improvement required to make imaging these granules a practical proposition is again within reach of SI. There are other interesting structures in this size range, such as the microvilli found on the surface of highly absorbent cells in the intestinal tract—hair-like extrusions that serve to greatly increase the cell surface area and can be destroyed by gastrointestinal infections. Even some viruses that are beyond current reach would be visible if the effective resolution limit were improved by a factor of two.

Another potential use for SI is as an adjunct to completely different techniques of biological investigation such as *scanning ion conductance microscopy* (SICM). This is a relatively slow but very high resolution process that measures the external topography of a cell by scanning a very fine pipette across its surface, adjusting position to maintain a constant conductance. Since the latter is characteristic of distance from the very high resistance cell membrane, it is possible to determine the physical shape of the cell—with a degree of accuracy well in excess of the optical resolution limit.

However, shape information alone is of limited use; ideally, it needs to be combined with information concerning *function*. The latter information is exactly what fluorescent probes are good at providing, within the resolution limits of optical microscopy. In order to benefit from the combination of both imaging techniques, the two must be performed simultaneously and the

results from both connected by appropriate image registration. The value of this registration is diminished if the lateral resolution of the fluorescent image is markedly lower than the SICM map—we will be matching up fine details with vague blurs. The improved resolution possible with SI can bring the two images into closer correspondence.

SI also has the virtue of providing a widefield image—one that is created *in toto* rather than requiring the sample to be mechanically scanned. The traditional alternative—which is, again, scanning confocal microscopy—can cause problems when combined with SICM, because the oil immersion objective (positioned underneath the sample) causes distortion of the cover slip as it moves, which in turn distorts the measurements of the simultaneously-scanning SICM pipette. The widefield SI process avoids having to move the objective relative to the sample, obviating the distortion.

These applications are only the tip of the iceberg. There is little doubt that a successful, fast SI implementation would be a boon to many researchers.

3 Microscope optics & the diffraction limit

3.1 Background

Although the modern microscope is a complex piece of high-precision optical technology, its basic structure is not conceptually very different from the pre-Enlightenment devices of Galileo and Huygens. The same fundamental physical constraints apply, the difference being that no 17th century microscope was ever in any serious danger of testing those constraints.

The microscope's purpose is to allow close inspection of very tiny objects far beyond the visual reach of the naked eye. To do so, a powerful, positive *objective lens* is used to produce a real *intermediate image* of the target, which is then magnified further by a second *eyepiece* lens to deliver an image to the eye. The combination of lenses effectively relocates an extreme close-up view of the object to a distance that our visual system can apprehend [16].

Since the initial light capture is performed by the objective, it is usually this lens that places the key constraints on what a microscope can resolve—subsequent stages of the process cannot recreate information that is absent at this point. Thus, the attributes of the objective determine the resolution of the microscope as a whole. This relationship may be characterised in a number of different ways; perhaps the most intuitive is due to Abbe, after whom its limit is named [3].

If we consider the spatial details of an object in Fourier terms, as a superposition of harmonic signals of different frequencies, then the resolvability of each detail will depend on the frequency of the periodic patterns that contribute to it. The smaller the detail, the higher the frequencies required for its representation.

Those frequencies are conveyed to the objective by waves of light, which undergo diffraction at the object as a result of the detail size; the eventual image is formed by constructive and destructive interference between these diffracted waves. For that to occur, at least the first order of diffraction from each detail must actually be captured by the objective. Finer details cause greater diffraction spread, which at some point will be beyond the range of the objective. So the resolving power depends essentially on the objective's breadth or *aperture*. Where the first diffraction order is wider than this, and thus not captured by the lens, the corresponding spatial details are not imaged (Figure 1).

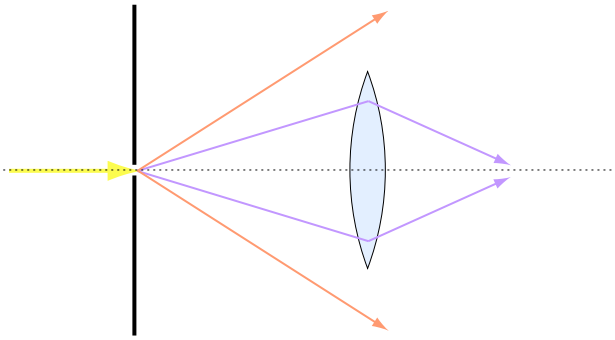

Figure 1: Incident light undergoes diffraction at an object. If a lens can capture at least the first diffraction order, an image can be formed. For smaller details, the first diffraction order escapes and no image results.

The degree of spread depends also on the refractive index of the medium between the sample (which we can assume for simplicity to be under a glass cover slip) and the objective. Although this will most often be air, some higher-magnification objectives are designed instead to be used with a denser medium such as water or oil. In these cases, the spread is somewhat reduced and the objective will capture a greater diffraction range.

These two resolution-governing properties can be conveniently combined into a standard dimensionless measure of detail capturing power called the *numerical aperture* or *NA*:

$$NA = n \sin \alpha$$

where n is the refractive index and α is the half-angle subtended by the objective at the focal point [3, 16].

In addition to the objective NA, a microscope's resolving power also depends on the wavelength of the light used to form the image: the effects of diffraction depend on the detail size relative to the wavelength, so longer wavelengths diffract more. Thus, the minimum resolvable detail size, d_0, is proportional to the wavelength, λ, and inversely proportional to the NA:

$$d_0 \propto \frac{\lambda}{NA}$$

It is sometimes more convenient to consider the maximum resolvable *spatial frequency*, k_0, which is proportional to the reciprocal of this value:

$$k_0 \propto \frac{NA}{\lambda}$$

In either formulation, the constant of proportionality depends on other aspects of the observational system, such as the degree of coherence in the light with which the image is being formed. A number of different values are commonly used according to context. A lower bound is set by the Abbe argument at

$$d_0 < \frac{\lambda}{2NA}$$

since at this point the first diffraction order for a marginal object is at the opposite edge of the aperture. But detail contrast diminishes to nothing as this condition is approached, making the practical resolution limit somewhat larger. A more useful value is given by the *Rayleigh criterion*, which considers the separability of object points in terms of their degraded counterparts—known as *Airy discs*—in the image plane. The smallest resolvable distance is held to be that at which the central maximum of one Airy disc intersects the first intensity minimum of the other. For the case of incoherent illumination, which will be our concern here, this is given by:

$$d_0 \approx 0.61 \frac{\lambda}{NA} \tag{1}$$

In addition to governing the *lateral* resolution, the NA also affects the *depth of field*, which is to say the thickness of the slice through the specimen that is acceptably in focus. A narrow depth of field is usually desirable, because it allows observations to be localised in the axial direction. However, in normal *widefield* operation, light originating away from the focal plane is still collected, polluting the image and diminishing the contrast of those details that are in focus.

The most popular technique for eliminating this stray illumination is *scanning confocal microscopy*: light is collected from only a single object point at a time, and a small *pinhole* aperture at the intermediate image position blocks rays emanating from out of focus sources [33]. Confocal microscopy is effective at creating an *optical section*, but the process is rather elaborate, as the image must be built up point by point in a raster scan over the sample. As well as requiring a good deal of expensive equipment, this also makes very inefficient use of light—and in particular fluorescent emission light—from the sample, the majority of which is simply discarded. Although light is in a sense *cheap*, it does nevertheless carry a cost on account of the fragility of fluorophores: the excitation energy that causes them to fluoresce can also denature them, leading to *photo-bleaching*: degradation into non-fluorescent compounds that no longer provide useful emissions.

3.2 Transfer functions

The majority of optical systems in common use are *linear*, meaning that the contribution of each element of the source object to the final image is independent of the others; the whole image can be taken as the sum of its parts [12]. It is often convenient to consider the image in these terms when characterising a lens's ability to convey information.

For any single point source in object space there will be a corresponding distribution of light in the image [29]. The mapping from one to the other is called the *point spread function* (PSF), $P(\mathbf{r})$, where $\mathbf{r}$ is a vector defining a point in image space.[3] For an object considered as a distribution of point intensities $O(\mathbf{r})$, the corresponding image, $I(\mathbf{r})$ will be the sum (or integral) of the point spreads of each:

$$I(\mathbf{r}) = \int_{-\infty}^{\infty} O(\mathbf{s})\, P(\mathbf{r}-\mathbf{s})\, d\mathbf{s}$$

This relationship is a *convolution*, which we denote with the operator symbol $\otimes$:

[3] We will assume for notational simplicity that image space and object space are isomorphic and use $\mathbf{r}$ to represent conjugate positions in both.

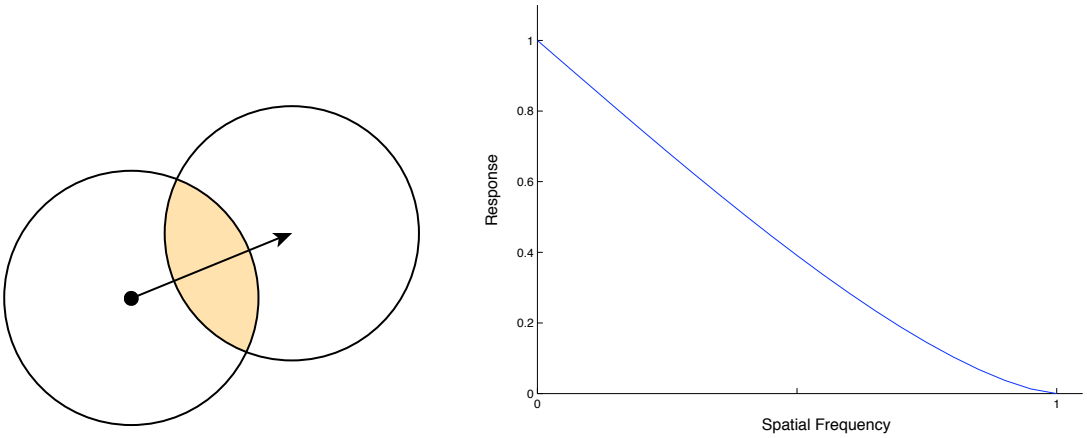

Figure 2: The OTF of an aberration-free lens system is equal to the proportional overlap of the pupil function with a displaced copy of itself (left). In the common case of a circular pupil, the OTF is circularly symmetric, with the frequency response in all directions declining smoothly to the cutoff value k_0.

$$I(\mathbf{r}) = O(\mathbf{r}) \otimes P(\mathbf{r}) \tag{2}$$

According to the well-known *convolution theorem*, the convolution of two functions in the spatial domain is equivalent to their product in the frequency domain (and vice versa):

$$\mathcal{F}\{g \otimes h\} = \mathcal{F}\{g\}\,\mathcal{F}\{h\} \tag{3}$$

where $\mathcal{F}$ denotes the Fourier transform, an operation that maps from one domain to the other [4].

Since multiplication is a more tractable operation than convolution, it is often convenient to use the transform of the PSF, known as the *optical transfer function* (OTF), $H(\mathbf{k}) = \mathcal{F}\{P(\mathbf{r})\}$. The OTF defines the *frequency response* of a lens system—that is, how well it transmits information of different spatial frequencies. In these terms, the image function becomes:

$$I(\mathbf{r}) = \mathcal{F}^{-1}\Big\{\, H(\mathbf{k})\,\mathcal{F}\{O(\mathbf{r})\}\,\Big\} \tag{4}$$

In the absence of aberrations, the OTF is equal to the proportional overlap of the lens pupil with a copy of itself displaced according to the frequency (Figure 2). At high frequencies, the pupils do not intersect and no information is transferred. In effect, this is just a restatement of the Abbe argument above, with the same limiting value.

In real systems there will always be some degree of aberration, which can be proved never to increase the modulus of the OTF [12]. This means that it cannot improve the contrast for any spatial frequency, confirming that a real lens must always fall short of the theoretical resolution limit.

3.3 Physical structure

The classical microscope has its optical components fixed within a tube looking down on the specimen stage [16]. Focus is adjusted by moving the entire unit relative to the stage (or vice versa). Optical microscopes for modern experimental biology typically use instead an *inverted* arrangement such that the objective is placed beneath the specimen looking up. The viewer still observes from above, the imaging path being bent by reflecting prisms. There are several reasons for this arrangement, notably that biological specimens typically reside in a liquid that the objective is not designed to be in contact with but which needs to be accessed by other experimental apparatus such as patch clamp pipettes. Placing the objective below allows a firm barrier, such as the underside of the slide or petri dish, to be kept between it and the sample without obstructing access from above.

Two distinct illumination pathways are employed. The standard 'brightfield' pathway illuminates from above with visible white light that passes through the sample to the objective. The specimen is viewed against a bright background. Although some viewing techniques require special filtering or polarisation, in most cases the image is illuminated and viewed directly. Moreover, the light source used is usually incandescent and its brightness can be turned up and down to set appropriate illumination levels. A standard condenser arrangement is sufficient for this case.

For fluorescence microscopy, the raw illumination source is heavily filtered to limit the excitation frequencies to a range appropriate for the chosen fluorophores, and the emissions are filtered again to protect the viewer's eyes from the potentially damaging excitation light and restrict the image to emissions of interest. The overall light levels are thus much lower, and the illumination must be suitably concentrated. Given the potential for photo-bleaching and the danger from stray excitation, it is further desirable to limit the illumination spot to the region being viewed. The solution to both problems is to use the microscope objective as a condenser, focussing the excitation field onto the sample at its own focal plane.

Such an arrangement requires the insertion of additional optical elements into the microscope's imaging pathway. In the classical microscope arrangement it is difficult to do this without disrupting the carefully-engineered optics and losing image quality. Instead, the structure in modern microscopes is modified to separate the light-gathering and image forming tasks of the objective, introducing an intermediate stretch over which the light is parallel. This region is known as the *infinity tube*—on the basis that the light within is focussed at infinity—and microscopes using such a configuration—as virtually all do nowadays—said to employ infinity tube optics (or sometimes just infinity optics) [27]. The lens taking on the image formation responsibilities is called the *tube lens* and is a fixed part of the microscope: it is not switched when changing objectives. The nominal magnification ($10\times$, $20\times$, etc) of the objective is actually dependent on the focal length of this separate lens (Figure 3).

One consequence of the more sophisticated optics and multi-purpose nature of such modern microscopes is that they are very restricted in terms of physical space. All the different elements must be kept in their relative places by a hefty metal superstructure; the various pathways are tied up into an intricate knot. As will be discussed in §7, this presents some difficulties when trying to add yet more functionality that was not envisaged by their designers, as is necessary for structured illumination.

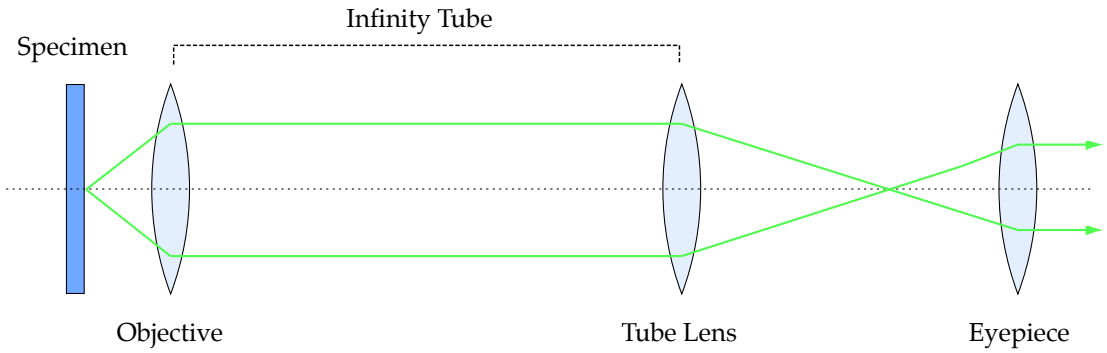

Figure 3: Infinity tube microscopes introduce a region of parallel light into the imaging pathway, which simplifies the addition of other optical elements.

4 Structured illumination

Although the diffraction limit is implacable, it does not represent a global embargo on the transfer of *information*, only a local one on the transfer of spatial frequencies. The idea underlying SI is that information *can* be transferred from regions of the object frequency space that are normally inaccessible to the OTF, provided it is first *shifted* into the accessible region.

The latter is achieved by introducing periodic signals into the excitation light, as will be described below; but that in turn leads to a good deal of other object information being lost. Consequently, the enhanced resolution image cannot be transferred in one go: it has to be smuggled across the border *in pieces*. Multiple frames with shifted excitation patterns must be recorded in order that the process can be inverted and something approximating the original signal—including otherwise inaccessible higher frequency components—reconstructed.

The techniques for improving resolution axially and laterally are somewhat different, and we shall consider them separately, beginning with axial resolution, which is one-dimensional and thus simpler. The question of managing both improvements simultaneously will be addressed in a later section.

4.1 Depth sectioning

As discussed in §2, a significant problem in widefield fluorescence microscopy is that the intensity of emission light does not sufficiently attenuate with defocus. As a result, the visible contrast of the object plane is reduced by stray light from elsewhere, making it difficult to localise our observations in the z direction.

What we would like to be able to do is separate light emanating from the plane of interest from that whose source is elsewhere. Confocal microscopy, described in §3.1, achieves this by erecting a physical barrier through which almost none of the out-of-plane light can pass. The cost is simultaneously to block the bulk of light from the object plane, permitting only a tiny point to be observed at a time. The image must be built up from a very large number of separate measurements [33].

SI provides for a quasi-widefield alternative by varying the excitation pattern in such a way that in focus emissions *can* be distinguished from out of focus. This is possible because only the zero order signal—which effectively makes up the unfocussed component—fails to attenuate.

The first and higher orders, being divergent, fall off rapidly, and this observation applies to the excitation light just as it does to the emissions [25].

If the subject is illuminated with a periodic pattern, then as we move away from the focal plane the pattern will go out of focus and become effectively uniform. If the pattern is shifted laterally, fluorophores in the plane will be excited or not according to the pattern's phase, while those away from it will be excited independently of the phase. Hence, images captured at different phases will contain a variable component from the focal plane and a constant component from the unfocussed volume. The latter can be eliminated algebraically, leaving only an axial section whose thickness depends on the illumination pattern frequency.

Although a geometrical optics model of the sectioning power can be readily visualised, this turns out to be a poor characterisation. Neil *et al* [25] develop an alternative approximation based on a model of defocus by Stokseth [32], and show that the axial intensity transfer is estimated by

$$I(u) \;=\; \left| (1 - 0.69\tilde{\nu} + 0.0076\tilde{\nu}^2 + 0.043\tilde{\nu}^3)\, \frac{2\, J_1[2u\tilde{\nu}(1 - \frac{\tilde{\nu}}{2})]}{2u\tilde{\nu}(1 - \frac{\tilde{\nu}}{2})} \right| \tag{5}$$

where u measures defocus and $\tilde{\nu}$ the spatial frequency of the imposed pattern at the sample. Both values are normalised relative to the numerical aperture and excitation wavelength, the relations being $u = \frac{8\pi}{\lambda} \sin^2(\frac{\alpha}{2})\, z$, for axial distance z, and $\tilde{\nu} = \frac{\lambda}{NA}\nu$, for actual spatial frequency ν. J_1 is the first order Bessel function of the first kind, while the multiplying polynomial in $\tilde{\nu}$ is Stokseth's empirically-fitted analytic approximation.

The Stokseth factor is independent of u, so it doesn't affect the sectioning strength, only the overall intensity range of the obtained image. The curve is shown in Figure 4; note that it falls to zero just before $\tilde{\nu} = 2$. Although this can to some extent be compensated by scaling, intensity resolution and SNR will suffer. Eventually, loss of contrast in the grating leads to frames that are indistinguishable and the calculated section becomes entirely black.

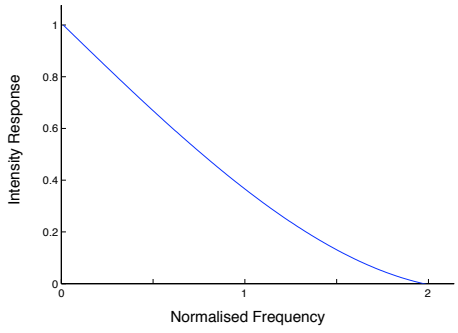

Figure 4: The Stokseth factor falls to zero as $\tilde{\nu}$ approaches 2.

The second part of the formula does depend on u and allows us to estimate the section thickness. As can be seen in Figure 5, attenuation occurs over the shortest distance where $\tilde{\nu} = 1$, while for the limit values 0 and 2—corresponding to uniform illumination and a spatial frequency at the Abbe limit respectively—there is no sectioning effect at all.

For simplicity, we define the section thickness as the width of the main peak of the intensity transfer curve. Table 1 gives this thickness in normalised units for a range of values of $\tilde{\nu}$, along with the corresponding thickness and grating pitch calculated for a 0.95 NA objective

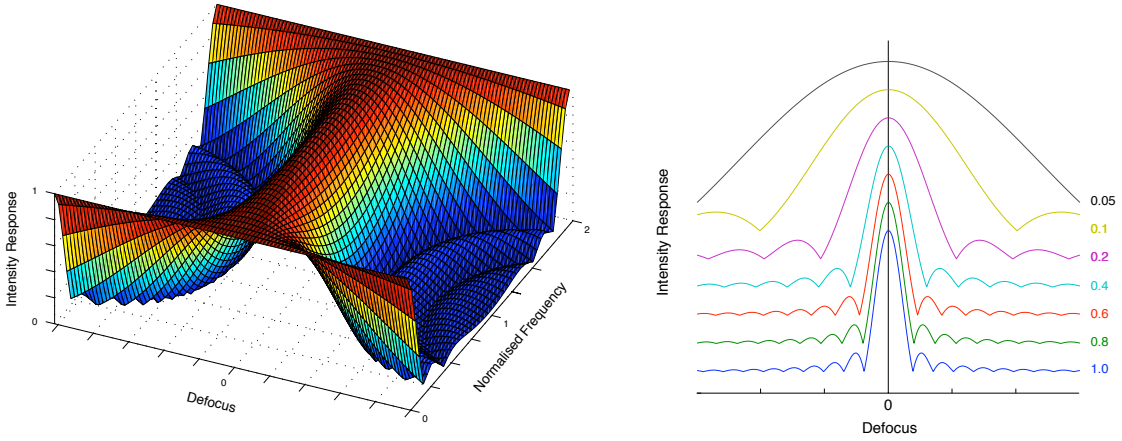

Figure 5: Attenuation with defocus is maximised when the normalised spatial frequency ($\tilde{\nu}$) of the excitation pattern is 1.

and 490nm excitation light. We observe that the thickness grows slowly at first with decreasing grating frequency: it is not necessary to generate exactly the optimal illumination pattern to see a benefit.

Normalised frequency $\tilde{\nu}$	**Normalised section**	**Grating pitch (nm)**	**Section thickness (nm)**
1	7.6	515	430
0.8	8.0	645	450
0.6	9.2	860	520
0.5	10.2	1030	580
0.4	12.0	1290	680
0.2	21.2	2580	1200
0.1	40.4	5150	2300
0.05	78.6	10300	4500
0.02	193.6	25750	11000

Table 1: Optical section thickness by grating frequency.

4.2 Lateral resolution

In §3.2, equations 2 & 4, the relationship between object and image was expressed in terms of the PSF and OTF, on the basis of a function $O(\mathbf{r})$ that represents the emission structure of the object. We now extend this model to include the effects of structured illumination and show how it allows additional detail to be passed.

The amount of light emitted at each point in the object space depends on the fluorophore concentration at that point and the intensity of the excitation light incident upon it. This relationship is essentially multiplicative—greater excitation stimulates the fluorophores to emit more—so given an object structure $O(\mathbf{r})$ and an excitation pattern $X(\mathbf{r})$, the corresponding emission pattern $E(\mathbf{r})$ will be:

$$E(\mathbf{r}) = X(\mathbf{r})\, O(\mathbf{r}) \tag{6}$$

By the convolution theorem, the effect in the frequency domain is:

$$\mathcal{F}\{E(\mathbf{r})\} = \mathcal{F}\{X(\mathbf{r})\} \otimes \mathcal{F}\{O(\mathbf{r})\} \tag{7}$$

That is, the frequency spectrum of the emission pattern is that of the object convolved with that of the excitation pattern. Convolution can be considered as a modulation of one spectrum by every point in the other, which implies the image will include multiple copies of the object spectrum with their zero frequencies displaced to the positions of the excitation spectrum peaks (Figure 6). It should be evident that this includes higher frequency information shifted toward lower frequencies. Such information is transmissible by the OTF [14, 19].

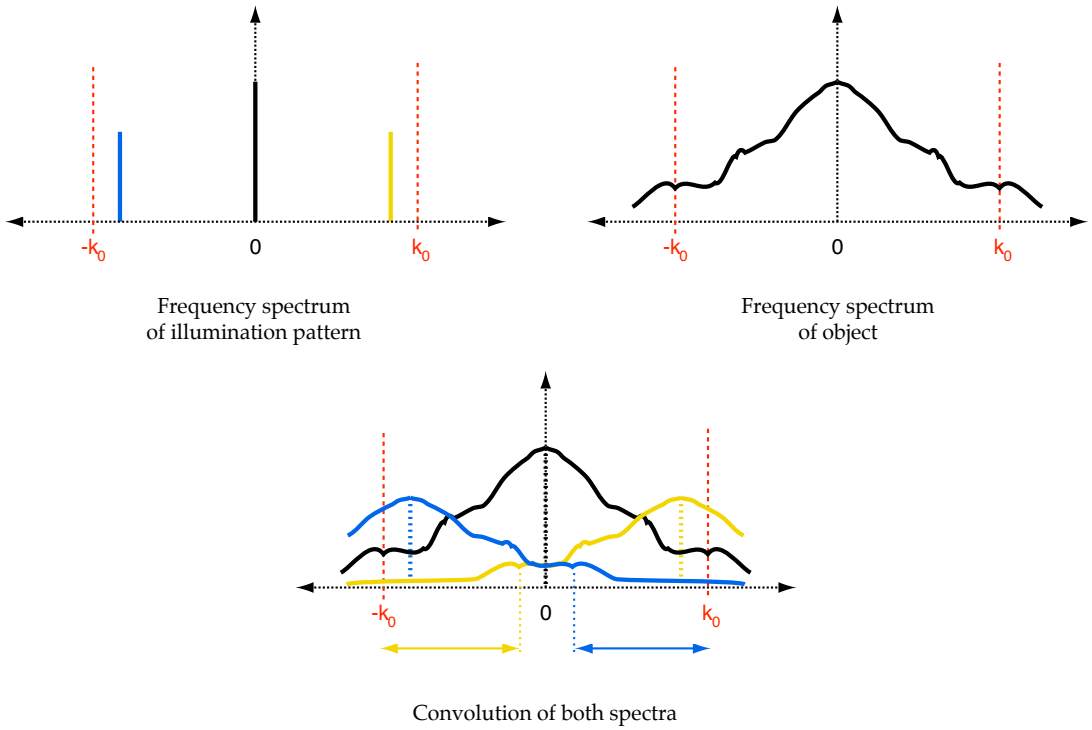

Figure 6: The convolution of the frequency spectrum of the illumination pattern and that of the object contains multiple copies of the latter scaled and displaced according to the peaks of the former. The indicated regions of the displaced spectra contain information from beyond the cutoff frequency k_0 shifted into the region of support. The observed spectrum will be the sum of those shown separately here.

This process can be readily understood in terms of *moiré fringes*, the familiar patterns we see in everyday life when periodic objects such as fabrics, grilles and chain-link fences overlap. These fringes are coarser than the objects that make them up, and can often be discerned even when the contributory patterns cannot [15].

Our high frequency information is, of course, not directly visible from the emission pattern, because it is mixed up with the other copies of the object spectrum. Moreover, a single image taken in this way only samples a fraction of the object space—the emission pattern contains no information from unexcited regions of the object. Once again, we must combine multiple images with phase-shifted illumination structure in order to recover an image of the object.

4.3 Pattern characteristics

Equations 6 & 7 conveniently gloss over the *nature* of the excitation pattern $X(\mathbf{r})$, and there are a number of choices we might wish to make. Important considerations include:

- Maximising resolution improvement
- Maximising sampling efficiency
- Minimising the computational cost of reconstruction
- Practicality of pattern generation

Although the last point will be more properly addressed in later sections, it should be clear that we require *some* mechanism to project the pattern onto the sample if we are to exploit this phenomenon at all. For the moment we will assume the use of incoherent light in a similar arrangement to that described for conventional epifluorescence microscopy in §3.3.

An important implication of this assumption is that the illumination pattern will be projected through a lens system that includes the microscope objective itself, and thus also subject to degradation by its OTF.

If we generate, by whatever means, some input excitation pattern $X_i(\mathbf{r})$ and then use our epifluorescence system to project it, the actual excitation field at the object, $X_o(\mathbf{r})$, will be the *image* of that pattern as formed by the illumination system, ie

$$X_o(\mathbf{r}) \;=\; X_i(\mathbf{r}) \otimes P_i(\mathbf{r}) \;=\; \mathcal{F}^{-1}\Big\{\, H_i(\mathbf{k})\, \mathcal{F}\{X_i(\mathbf{r})\} \,\Big\} \tag{8}$$

where P_i and H_i are the illumination PSF and OTF, respectively. Evidently, X_o will be limited by the passband of H_i. In the best case, this may be marginally better than the imaging passband, since the excitation light is at a shorter wavelength, but the imaging OTF represents a good first approximation to the limit. (As will be discussed later, the effective numerical aperture of the illumination pathway may actually be considerably worse than that of the objective alone, with a correspondingly tighter cutoff.)

As can be seen in Figure 6, the range of frequencies that are displaced into the accessible region depends on the frequency of the peak to which they are attached. For an imposed spatial frequency $\mathbf{k}_1$, a circular region of frequency space centred on $\mathbf{k}_1$ of radius k_0 is translated into the accessible region, effectively extending that region (in the direction of $\mathbf{k}_1$) to a radius $k_0 + |\mathbf{k}_1|$. Since, from the above argument, $|\mathbf{k}_1| \leq k_0$, the cutoff frequency—and hence the lateral resolution—can potentially be doubled [14].

Now, consider the properties of the excitation pattern in the context of the considerations listed earlier. We see immediately that lateral resolution will be maximised if the pattern contains frequencies as close to k_0 as possible. However, this is at odds with the axial sectioning requirement outlined in §4.1, for which the optimum illumination frequency $\tilde{\nu} = 1$ corresponds to $\frac{1}{2}k_0$, while a frequency actually at k_0 offers no sectioning power at all. There is thus a tension between these different goals that must be resolved according to the particular experimental priorities. The choice is not strictly binary: even the optimal sectioning pattern will still lead to a 1.5× lateral resolution improvement, so there is some scope for compromise.

The criteria for maximising sampling efficiency and minimising reconstruction costs are closely related, and depend to some extent on the recovery algorithm to be used [18]; several of these will be discussed in the next section. For the most part what these criteria boil down to is taking

as few frame captures as possible while still being able to piece together a complete image. The more frames we need, the more our SI endeavours will sacrifice time resolution.

For an imposed one-dimensional periodic pattern, any single phase shift will inevitably leave some image points identically illuminated and hence provide insufficient information for full recovery. We therefore need a minimum of 3 captures for this case, and the optimal pattern will be a single-frequency sine grating [25].

Such a configuration is sufficient for depth sectioning and can also be used to provide lateral improvement in one direction. The latter is unlikely to be of much interest in itself, however, so in cases where we are aiming for better lateral resolution we will need to extend the improvements to two dimensions. This can be done either by using one dimensional patterns with different orientations or by using a two dimensional pattern with two dimensional phase shifts. The economics of the former are straightforward: for n orientations, we need to capture at least $3n$ frames; at a minimum, then, we need 6 frames for two perpendicular grating orientations.

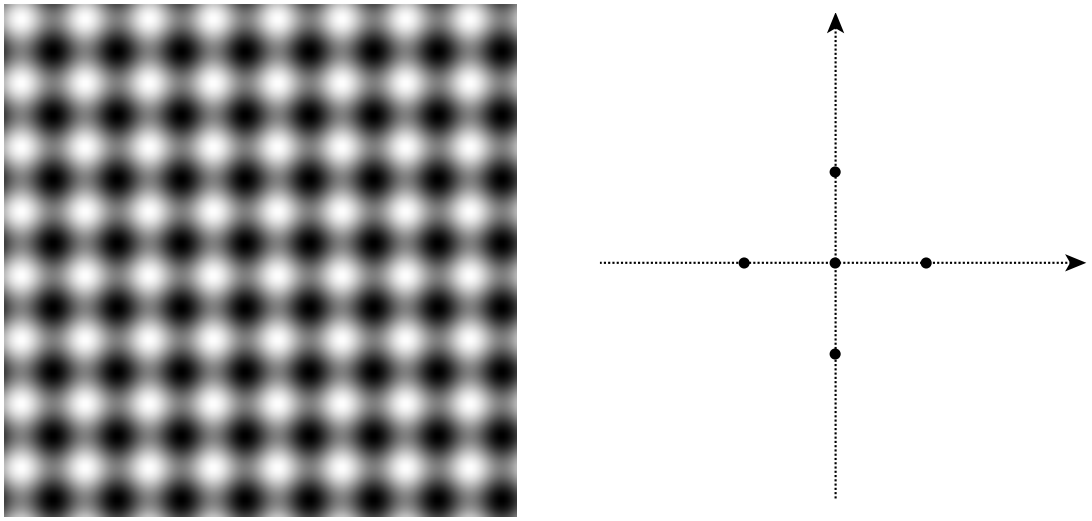

Figure 7: A simple 2d 'gingham' pattern, formed from the sum of two perpendicular sine gratings, and its corresponding frequency power spectrum.

With a two dimensional pattern, the situation is somewhat more complicated. Using the most sophisticated reconstruction algorithm, the minimum number of captures needed depends on the number of peaks in the pattern's Fourier spectrum [20]. A minimal 'gingham' pattern such as depicted in Figure 7 will have 5 peaks, so we should in theory be able to reconstruct an image from just 5 captures, given appropriate phase shifts in both directions.

However, there is—as so often—a trade-off between the efficiency of capture and the computational complexity of reconstruction. For the algorithm that works with this number of sample frames, reconstruction may not be possible in real time. Faster algorithms can be applied, but the images they produce will suffer from modulation artefacts. The choice is then to put up with real time imperfections and run the better reconstruction offline afterwards, or to expand the set of captured frames and generate results in real time using a different algorithm and a lower frame rate.

Once again, this choice will be governed by experimental priorities; and once again, the choice is not binary. A variety of compromise positions are possible. Our baseline recommendation for lateral resolution would be just such a compromise: a 5 frame capture regime, but using

only 3 for real time recovery and rendering via one of the cheaper methods. Provided all data are stored, full reconstruction can be performed *post hoc* for interesting sequences.

One final consideration as regards the illumination pattern is the level of contrast that can survive the OTF. This will affect the strength of the imposed signal, and in turn will strongly influence the recovered SNR. Two dimensional patterns include periodicities in pretty much all directions, all of which will be subject to attenuation during projection. The consequence is that these patterns will be transmitted less faithfully than their one dimensional counterparts, with more of the excitation energy being redistributed to the zero order. The resulting poor separation between light and dark regions will lead to amplification of noise in the recovered image. There may therefore be an argument for using a less efficient one dimensional illumination scheme in cases where the SNR is a concern—in particular, where the sample has low concentrations of fluorophore, or where things are moving so quickly that the frame exposure times must be kept very short.

4.4 Reconstruction

Given a set of captured frames with varied excitation patterns, a number of techniques may be employed to recover the improved image. While some are tied to a particular illumination strategy, others are more widely applicable. Each has its own characteristics in terms of robustness, computational complexity and the nature of the improvements obtained, so the best choice will depend on the context.

If the constituent frames' illumination patterns sample the object plane evenly, then a naïve reconstruction can be obtained by simply taking the sum of all frames [25]:

$$I = \sum_i I_i \tag{9}$$

The result is equivalent to a single frame taken with uniform illumination.[4] This discards any benefit from the patterning; to obtain that, account must be taken of the way the image at each point varies with the excitation pattern.

Our first goal is to eliminate the unvarying component and leave an axial section as discussed in §4.1.

Using one-dimensional sinusoidal illumination, this problem is equivalent to identifying the amplitude of the sine wave at each point. This can be done exactly provided we have at least three samples per period at known relative phases. For the case of even spacing—ie, with the grating shifted by $\frac{2\pi}{3}$ each frame—the calculation is as follows [25]:

$$I = \sqrt{(I_1 - I_2)^2 + (I_2 - I_3)^2 + (I_3 - I_1)^2} \tag{10}$$

This eliminates the constant component and maps the detected signal at each point into a uniform range. Provided the illumination patterns are consistent with the deriving assumptions, the resulting image should be free of modulation artefacts. For reasons discussed in §4.5 below, some effective lateral resolution improvement will also be obtained parallel to the line grating,

[4] There will generally be some constant factor difference according the number of frames and the amount of illumination in each frame. We assume throughout this section that all calculations are performed with sufficient numerical precision and the reconstructed image undergoes a final normalisation, so such constant factors will be ignored.

and this may be exploited by taking several sets at different grating angles and combining them additively.

The calculation is specific to a sinusoidal illumination pattern and fixed phase shifts. It does not easily generalise to doubly-periodic patterns and must be recast for a different set of phases. For example, the non-uniform shift set $\{0, \frac{\pi}{2}, \pi\}$, which offers some advantages when generating the pattern electronically, is instead recovered using the reduced formula [10]:

$$I = \sqrt{(I_1 - I_2)^2 + (I_2 - I_3)^2} \tag{11}$$

It's clear that this technique is sensitive to any vagaries in the pattern production process, which will lead to banding. Nevertheless, it provides clean and straightforward recovery. In the form expressed by equation 10 it is used in existing commercial implementations such as the Zeiss Apotome.

An alternative approach [18] determines the value at each position according only to the highest and lowest intensity recorded at that position over the complete set of frames:

$$I = \max_i(I_i) - \min_i(I_i) \tag{12}$$

Equivalently, this is the maximum intensity value across the set of difference images:

$$I = \max_{i,j}(I_i - I_j) \tag{13}$$

The rationale for this approach is intuitively simple: the maximum recorded for a point will occur when it is most strongly illuminated, the minimum when it is least strongly illuminated. In the latter case the signal will be mostly due to out of focus contributions, and we eliminate these by subtraction.

If we could be certain of sampling every object point in both fully-excited and fully-unexcited states, this calculation would indeed leave only the focal section. Unfortunately, due to the blurring effects of the PSF, this cannot be easily achieved for spatial frequencies close to the resolution limit, and there will usually be some degree of modulation due to the variation in illumination range.

Despite this susceptibility to modulation artefacts, this technique has the advantages of being computationally simple and extremely general: its only requirement of the illumination strategy is that the object be thoroughly sampled in both light and dark states. The algorithm is in large part independent of the choice of excitation pattern and number of frames.

The techniques discussed so far do not explicitly account for the Fourier domain convolution effect described in §4.2, concentrating instead on the removal of out of focus contributions; that they can nevertheless obtain apparent resolution benefits is something of a surprise, as will be discussed in §4.5. The remaining two approaches, on the other hand, are specifically designed to separate and shift the duplicated object spectra to recover a more detailed image.

A relatively simple such method appears in one of the earliest papers demonstrating practical lateral resolution improvement [20]. The form presented is tailored to one-dimensional sinusoidal illumination at phases $\{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. It exploits the linearity of the Fourier transform to perform the whole process in the spatial domain, and also uses the symmetry of the alternate phases to simplify the calculation.

The process is summarised by the following equation:

$$I \;=\; \mathrm{Re}\Big\{ \big[(I_1 - I_3) - i(I_2 - I_4)\big] e^{i(\mathbf{k}.\mathbf{r}+\phi_0)} \Big\} \tag{14}$$

where $\mathbf{r}$ is the position vector in real space and $\mathbf{k}$ represents the grating frequency and direction (which is constant over the frame set). The factor $e^{i(\mathbf{k}.\mathbf{r}+\phi_0)}$ is thus a one-dimensional complex sine grating of the same frequency as the illumination pattern; its function is to shift the displaced object spectrum back to the correct position in Fourier space. The phase angle ϕ_0 is chosen such that the relocated zero-frequency is real and positive.

As with any one-dimensional illumination strategy, the resolution benefits are *directional*, providing optimum improvement parallel to the imposed pattern, none at all perpendicular to it. For two-dimensional improvement, further frame sets with the grating at different orientations must be captured and combined.

A much more general approach to the same problem is to frame it in terms of linear algebra [14, 17]. In some cases this may still be implemented entirely in the spatial domain, but we present it here in a form that assumes we are willing to perform Fourier transforms as needed. While the method is not tied to any particular illumination strategy, it must be parameterised according to the Fourier structure of the excitation pattern used.

Given a pattern whose Fourier spectrum contains a known set of m peaks (limited by the passband of the illumination OTF), the detected image spectrum will consist of the convolution of the object emission spectrum with those peaks. Each sampled Fourier point therefore consists of the sum of several different object spectrum points weighted according to the peak to which they are attached, with the result attenuated by the OTF:

$$\tilde{I}(\mathbf{k}) \;=\; H(\mathbf{k}) \sum_{i=1}^{m} \gamma_i \, \tilde{O}(\mathbf{k} - \mathbf{k}_i) \tag{15}$$

The tildes here are a reminder that we are using the frequency domain versions of the image and object functions, the $\mathbf{k}_i$ are the illumination peak positions in Fourier space and the γ_i the corresponding complex peak sizes. Each image frequency point can thus be seen as an equation in m unknowns.

When the excitation pattern is shifted in the real domain, only the coefficients γ_i change—the peaks themselves do not move [4]. A set of n such shift images thus provides a system of simultaneous equations at each $\mathbf{k}$ that can be solved to estimate m points from the wider object spectrum:

$$H(\mathbf{k}) \begin{pmatrix} \gamma_{1,1} & \cdots & \gamma_{1,m} \\ \vdots & \ddots & \vdots \\ \gamma_{n,1} & \cdots & \gamma_{n,m} \end{pmatrix} \begin{pmatrix} \tilde{O}(\mathbf{k} - \mathbf{k}_1) \\ \vdots \\ \tilde{O}(\mathbf{k} - \mathbf{k}_m) \end{pmatrix} = \begin{pmatrix} \tilde{I}_1(\mathbf{k}) \\ \vdots \\ \tilde{I}_n(\mathbf{k}) \end{pmatrix} \tag{16}$$

Note that for almost all $\mathbf{k} - \mathbf{k}_i$ in the expanded region of support there will be multiple equation systems producing an estimate, each affected differently by the OTF. If a reasonable model of the OTF is available, it can be used to produce a weighted average to regularise the spectrum; this should also compensate for sensor noise [17]. Otherwise, the combined spectrum will give rise to artefacts, though these may be reduced by a final apodisation [15].

Given that the calculation depends on knowledge of the illumination spectrum peaks, the problem remains of actually identifying the $\mathbf{k}_i$ and $\gamma_{j,i}$. *A priori* prediction from knowledge

of the pattern is almost certainly unreliable, especially when it comes to locating it within the captured frame. A better alternative is to identify the peaks from the maxima of the cross-correlation function of the image [17].

This reconstruction technique is considerably more sophisticated and flexible than those based on direct image processing alone, and can be generalised further to make use of non-linear emissions (for example, due to saturation of the fluorophores), permitting lateral resolution improvements in excess of the $2k_0$ limit we face when using a linear response. However, the computational cost is significant, and the number of frames needed is also large.[5]

4.5 Edge enhancement, lateral resolution & aliasing

The reconstruction methods represented by equations 10-14 in the previous section all depend on combining differences between pairs of real images. As well as eliminating the zero order, this has the side effect of enhancing edge-like structures perpendicular to the periodicities in the illumination pattern, leading to a decrease in the minimum resolvable separation distance. Somewhat paradoxically, this effect is a consequence of the PSF itself, which leads each detected image point to contain some information about neighbouring object points.

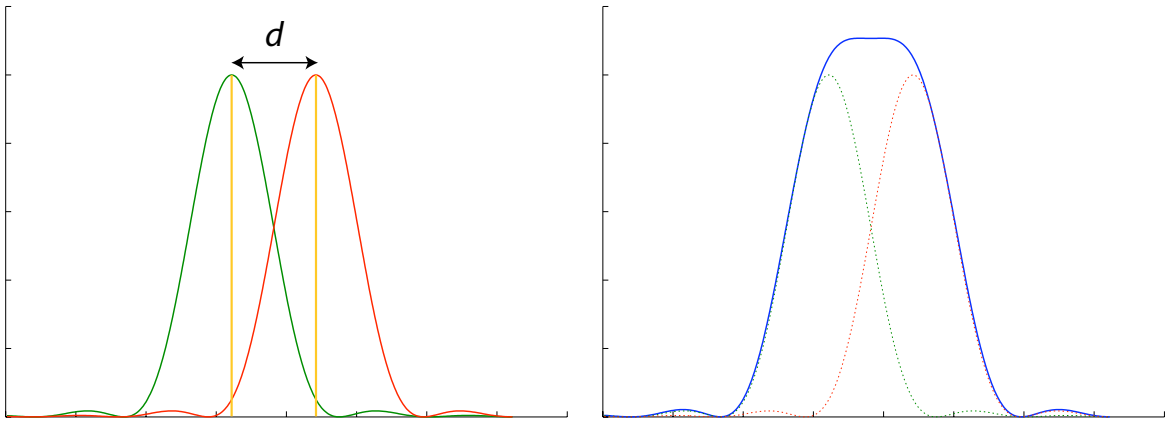

Figure 8: Nearby emission points coalesce under the PSF.

Consider the one-dimensional case of two single-point concentrations of fluorophore at a separation d, as in Figure 8. When excited uniformly, each emits the same illumination intensity and the image obtained is the convolution of those two points with the PSF (depicted as an Airy pattern, $(J_1(r)/r)^2$). As predicted by the Rayleigh criterion, for $d \leq d_0$ the points coalesce and cannot be resolved.

If instead the points are illuminated by a sine grating of spatial frequency ν at different phases ϕ_i, then—provided the separation is not an integer multiple of the grating period—the *relative* intensity of emitted light from the two points will vary with phase:

$$I_i(r) = P(r)\sin(\phi_i) + P(r-d)\sin(\phi_i + 2\pi\nu d) \tag{17}$$

[5] At first glance it would seem that there must always be at least as many captures as peaks for the equation system to be non-singular, ie $n \geq m$. In fact, it is sometimes possible [20] to exploit symmetries in the peak structure in order to construct an invertible $n \times n$ system with $n < m$. This only really becomes relevant in the non-linear case, though, when n is already comparatively large.

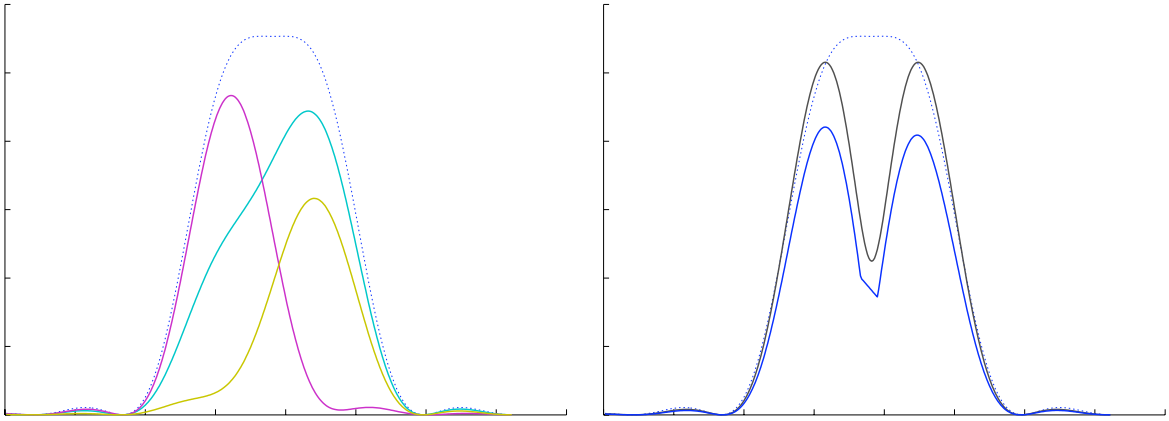

Figure 9: Varying skew in the emissions at different SI phases (left) allows objects to be 'inadvertently' resolved (right: upper curve is the Neil reconstruction, lower the max-min).

Since the points are too close to resolve, each image will still register only a single peak, but that peak will be skewed in the direction of the more brightly illuminated point, as seen in Figure 9. When the information from several phases is combined, the separate peaks can be detected. Note that the design of equations 10-13 takes no account of this spreading effect, being based only on the separation of the out of focus component. That an enhancement occurs at all is really an accident.

The apparent increase in resolution depends on the relationship between the periodicities in the object, the illumination period and phases, and the point spread function. Further, it is tied to the non-linearity of the reconstruction formulæ: a linear combination would be unable to introduce frequencies not already present in the contributing frames. To the best of our knowledge this effect has never been described analytically, and indeed, while we can make some observations as to what gives rise to it, we haven't achieved a proper characterisation within the timeframe of the project. We believe it would be beneficial to do so and suggest that as an obvious locus for future work.

Finally, we note that although this enhancement can genuinely improve resolution of some object details, it is nevertheless in some sense an *artefact* of the reconstruction process, and can in fact reduce fidelity even as it produces the semblance of greater resolution. It leads to dilation at object boundaries and distortion of tonal variations. The max-min procedure in particular is prone to introduce high-frequency aliasing that could be misleading if not filtered or at least recognised. While there are many biological applications for which this will not be an issue, the distortions must certainly be taken into account during analysis if the image is being generated for quantitative purposes [2].

4.6 Simulation results

To illustrate some of the features discussed above, simulations of image formation and reconstruction were performed using the MATLAB environment. Images were generated from a test object and illumination pattern according to this formula:

$$I = \mathcal{F}^{-1}\Big\{H\,\mathcal{F}\Big\{N + O\,\mathcal{F}^{-1}\left\{H\,\mathcal{F}\{X\}\right\}\Big\}\Big\} \tag{18}$$

where X is the generated excitation pattern, O the object, N a noise term representing out of

focus emissions and H the OTF. The latter was taken to have the ideal displaced-pupil form mentioned in §3.2, which for a circular pupil can be formulated as [12]

$$H(\mathbf{k}) = \frac{2}{\pi}\left[\cos^{-1}\left(\frac{|\mathbf{k}|}{k_0}\right) - \frac{|\mathbf{k}|}{k_0}\sqrt{1-\left(\frac{|\mathbf{k}|}{k_0}\right)^2}\right] \tag{19}$$

Although there is scope in the code for the OTF cutoff to differ between input and output—which will likely be the case in practice—our simulations were obtained with H equal for both, since this corresponds to the best we can reasonably expect from epi-illumination via the objective.

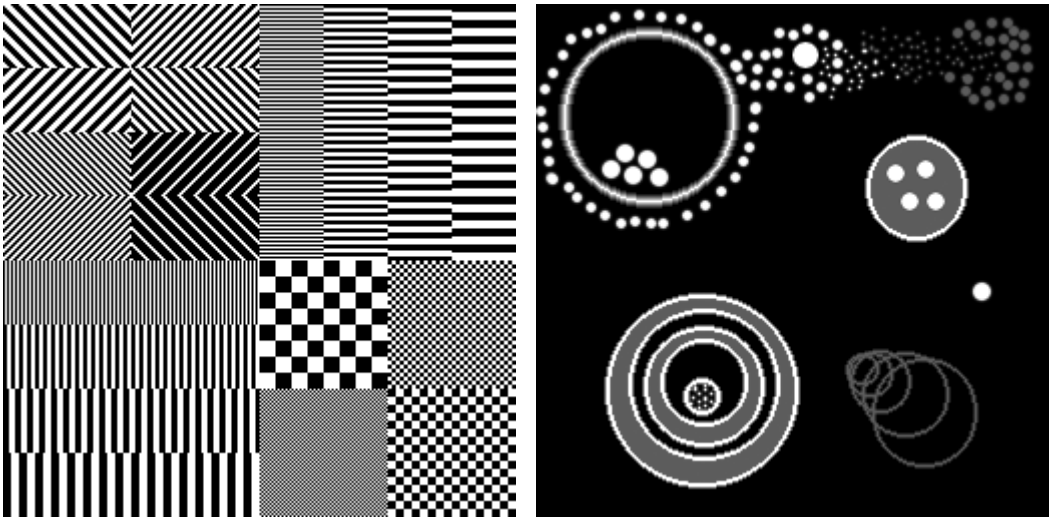

Figure 10: The target images used as objects for our simulations.

Two different target images were used, as shown in Figure 10. The first is highly geometric, intended to contain strong high-frequency components; the second to test the resolvability of features of more biological relevance, since resolution improvements will be of little value if they only work for rigidly rectilinear specimens. Examples of the corresponding image-side versions of each object under uniform illumination, $X = 1$, can be see in Figure 11. A clear loss of finer detail can be readily observed. (Here are and in subsequent images, a more restrictive OTF is applied to the second target than the first, to give a clearer impression of the effects.)

Sets of frames were then produced using phase-shifted illumination patterns, starting with a one-dimensional grating. Example frames for both objects are shown in Figure 12. Reconstructed images using the Neil method (equation 10) and the max-min method (equation 12) are shown in Figures 13 and 14 for a single dimension and Figures 15 and 16 for two perpendicular sets combined additively. We note several features of these reconstructions:

- The out of focus component appears to be successfully removed.
- A modulation at the same frequency as the imposed grating can be observed in the max-min case, although it is relatively weak in this example and the two-way addition effectively removes it. We would expect this to be more pronounced in flat areas of bright emissions, but such areas would be atypical of biological samples.
- Both methods show clear edge enhancement parallel to the grating, as described in §4.5. This is particularly obvious in the one-way case, where it leads to an asymmetric response

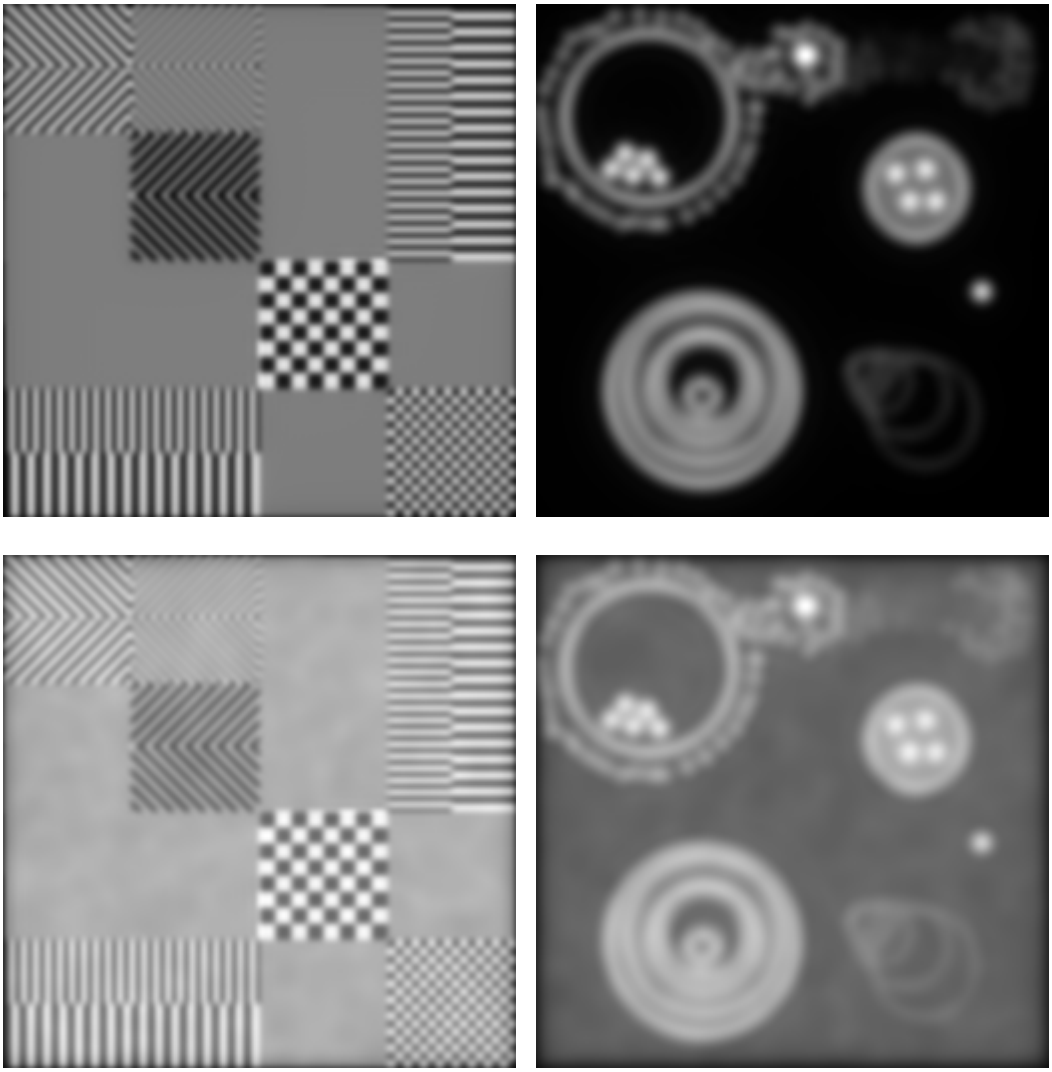

Figure 11: Simulation images under uniform illumination. Lower images include additive out-of-focus emission noise.

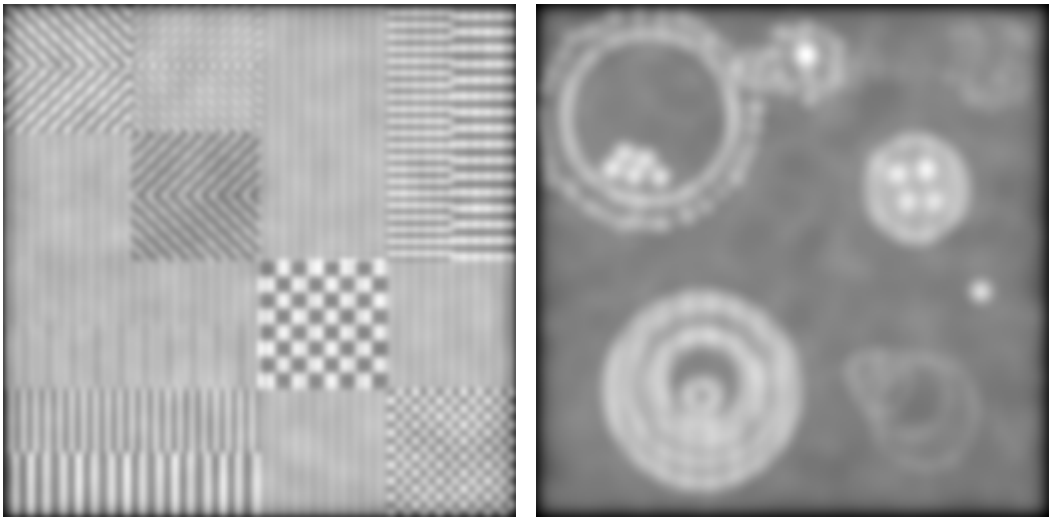

Figure 12: Single simulated frames with sine grating illumination.

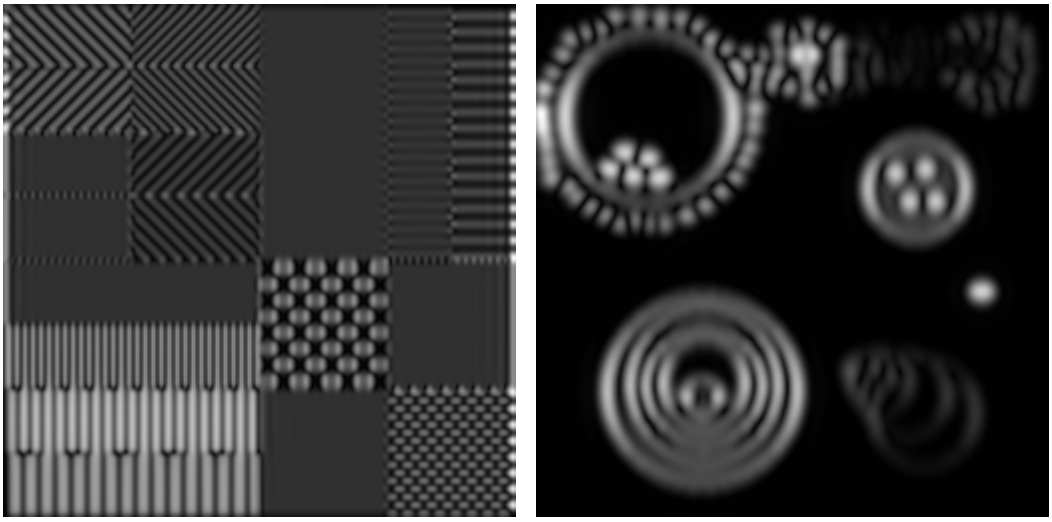

Figure 13: Images reconstructed from three frames by the Neil method.

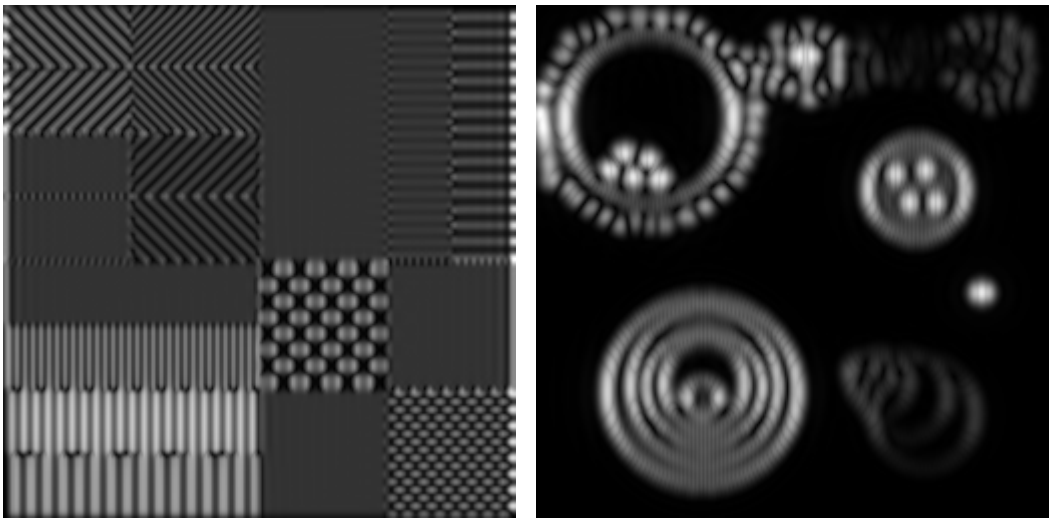

Figure 14: Images reconstructed from three frames by the max-min method.

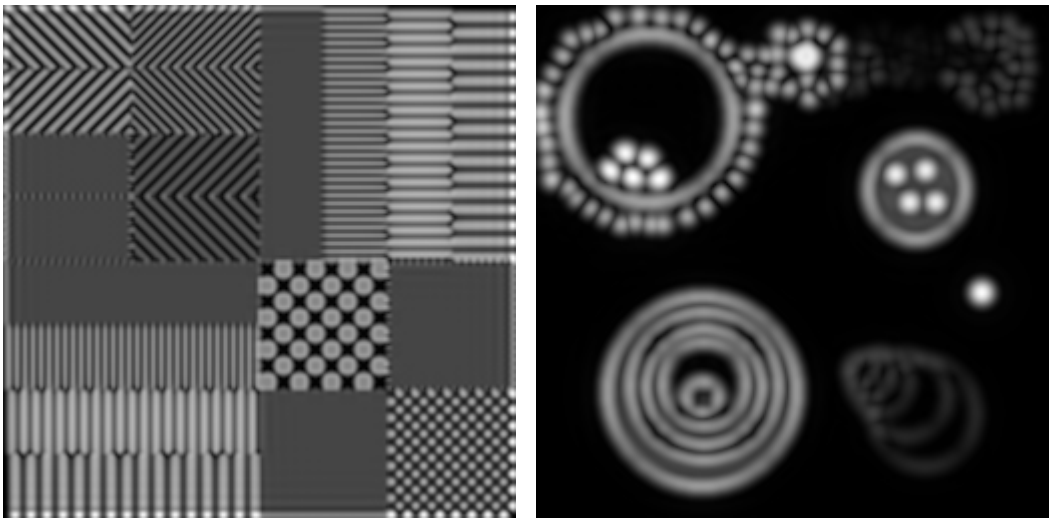

Figure 15: Additive combination of perpendicular Neil reconstructions.

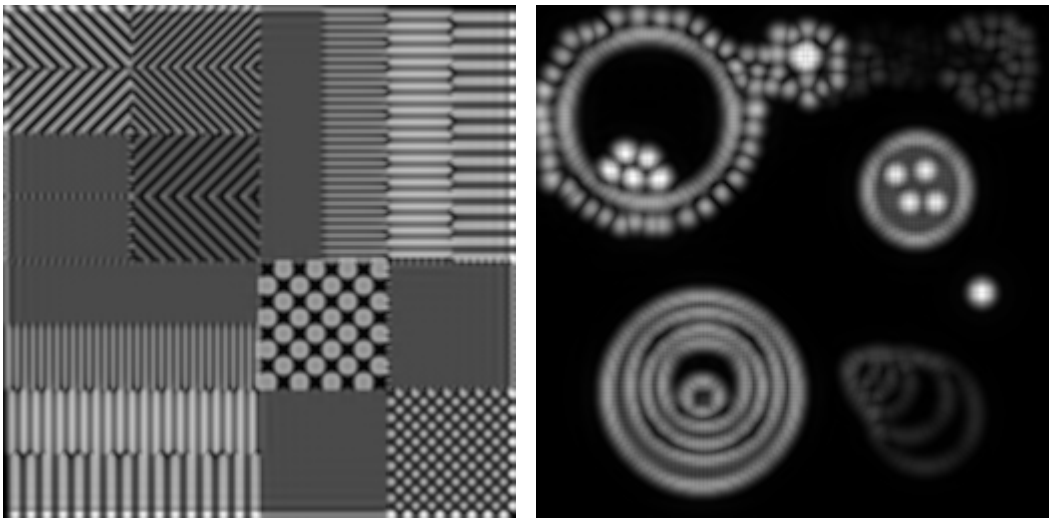

Figure 16: Additive combination of perpendicular max-min reconstructions.

with pronounced artefacts. The two-way additive version smooths these out somewhat by averaging the overstatements of each direction with the understatements of the other.

- Distinct high-frequency moiré artefacts can be observed in the max-min reconstructions at a level of detail well beyond that OTF's region of support. Once again, these are to some extent ameliorated by combining perpendicular versions, but they remain quite noticeable.

The Neil method does not generalise easily to illumination with a two-dimensional pattern, but max-min simulations were carried out using a checkerboard illumination pattern with a complete set of $\frac{1}{3}$-period shifts in both directions (for a total of nine sample frames). Figure 17 shows single example frames with such illumination, while the max-min reconstructions are in Figure 18. The resulting images have less exaggeration of the verticals and horizontals, with a more evenly-distributed edge enhancement. A small directional bias can be seen on the diagonals instead, which are the dominant spatial frequencies of the checkerboard, but it less prominent than in the additive sine grating version. High frequency moiré artefacts remain in evidence, though again they are slightly more even.

All the simulations so far have used a noise model N that is constant over the set of sample frames, consistent with the underlying assumptions of the depth sectioning procedure. In practice, however, there will also be some variable noise component inherent in the sampling process. Since all proposed reconstruction methods implicitly assume that the illumination pattern is the only variable across frames, such noise will impact the quality of the recovered image.

Although this problem has not been explored in detail, some simulations with variable noise were run to get at least some sense of how damaging it might be. Results can be seen in Figures 19 and 20. A simple Gaussian noise model was used, and in order to make the effect more visible frames were generated with no constant out of focus contribution. Even so, noise levels that are all but imperceptible in the sample frames give rise to significant degradation in the reconstructed image. While the impact of this effect will to some extent depend on the particular noise model, it is likely to present problems with all but the most noise-free equipment.

Some further simulation images can be found on the project website at `http://www.ucl.ac.`

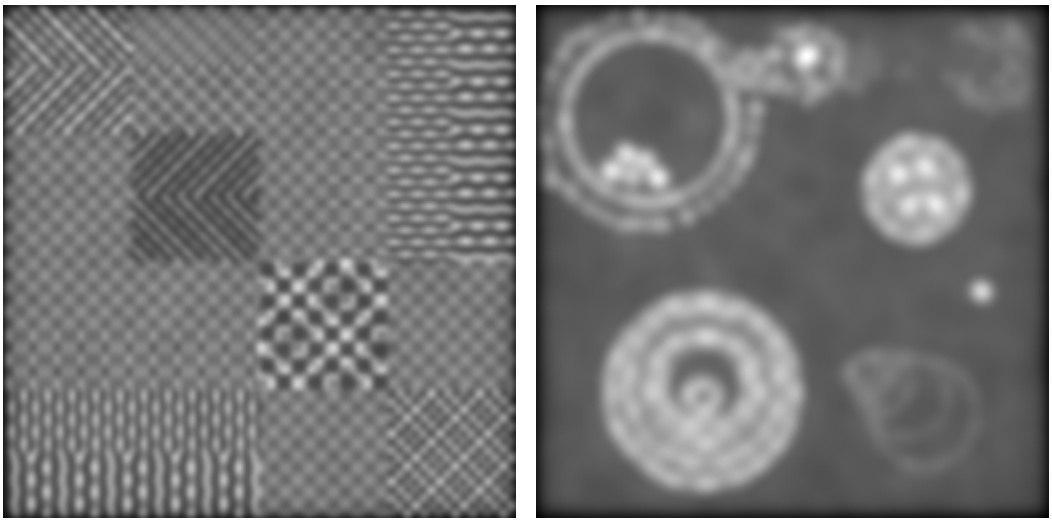

Figure 17: Single simulated frames with checkerboard illumination.

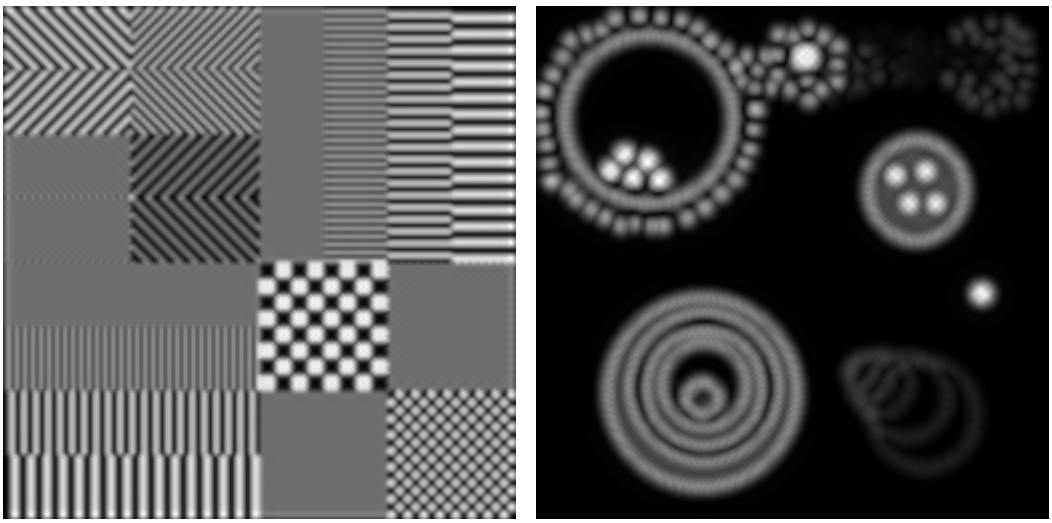

Figure 18: Max-min reconstructions from nine checkerboard-illuminated frames.

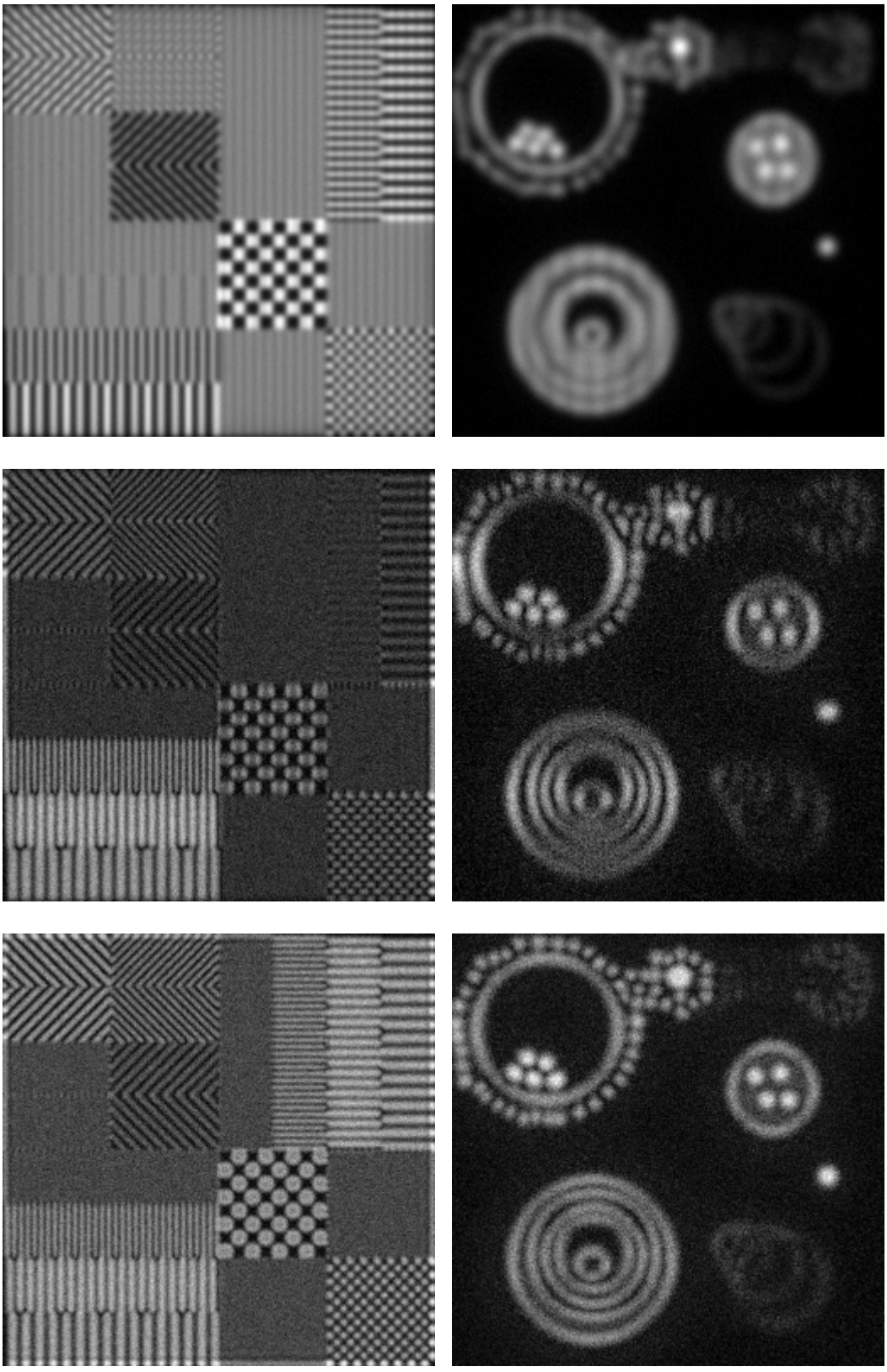

Figure 19: Reconstructions from data with zero-mean Gaussian noise, variance 0.00001. Single frames (top) show minimal ill-effects, but reconstruction amplifies the differences.

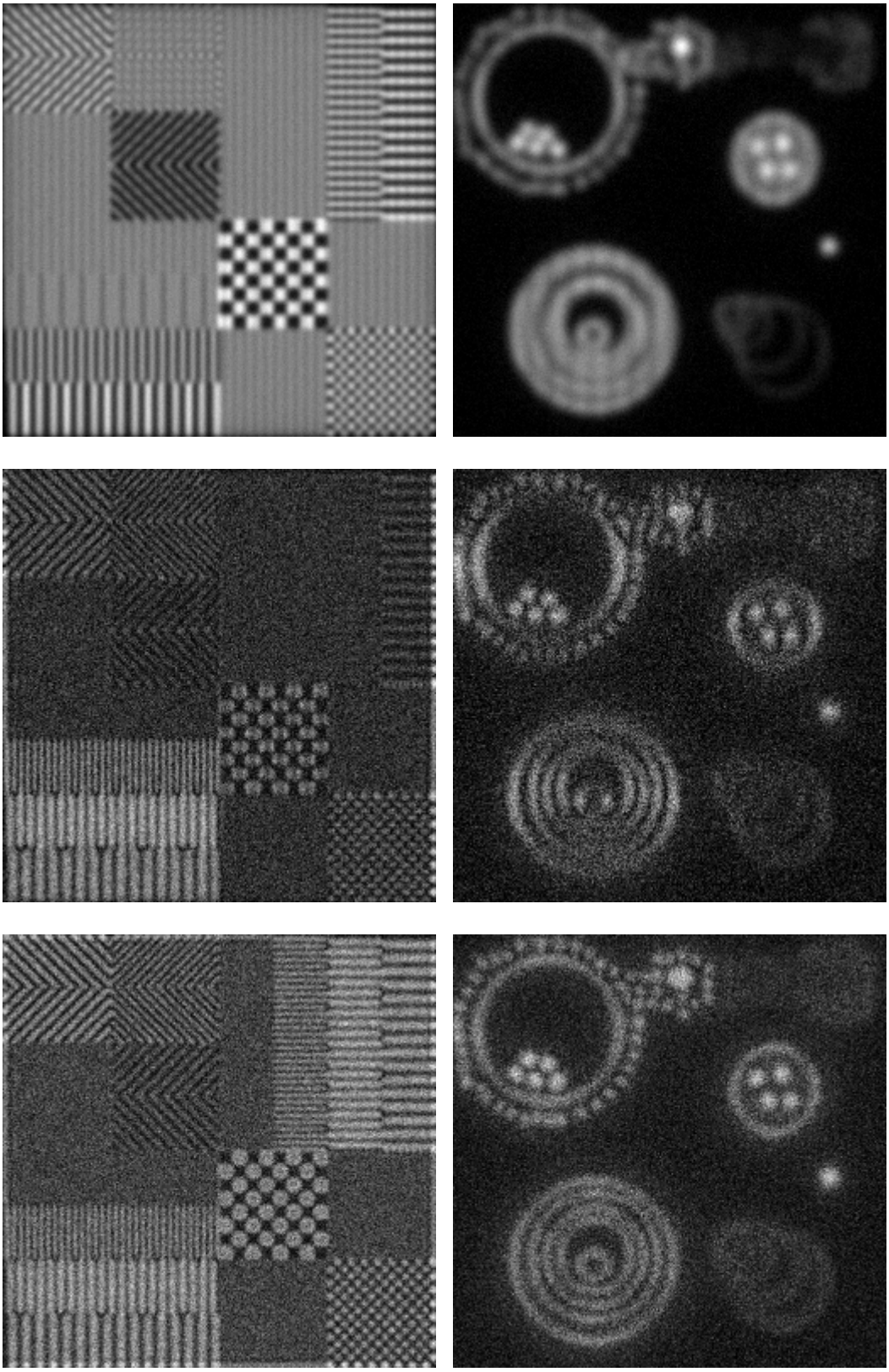

Figure 20: Reconstructions from data with zero-mean Gaussian noise, variance 0.00005.

uk/~ucbpmbc/project.html. Included are some animated sequences of multiple frames with shifted illumination patterns, which are helpful to illustrate the variations exploited by the reconstructive process.

Selected MATLAB source code for these simulations is reproduced in Appendix A and also available in electronic form at the above web page.

4.7 Limitations of current implementations

SI is a relatively recent innovation and is the subject of a number of active research projects. Most existing implementations are internal research builds of an inevitably DIY nature, but the depth sectioning benefits in particular have potentially very broad appeal and at least one major microscope manufacturer, Zeiss, markets a commercial implementation, the Apotome, as an add-on to its laboratory microscopes [8]. In common with most of the published research versions, this uses a glass diffraction grating to generate the illumination patterns.

Such a mechanism benefits from being relatively cheap, physically dependable and well understood, but it lacks flexibility and is constrained by the need to physically move it in order to shift the illumination pattern. This is practical for depth sectioning, which requires pattern shifts in only one direction, although it imposes a significant limit on the speeds that can be achieved, with the concomitant problems outlined in §2. There is also the inconvenient requirement to change the grating for different objectives, since its line pitch is, naturally, fixed. For lateral resolution improvements, which demand orientation changes as well, a glass grating is more or less useless for anything remotely real-time or biological.

Some research implementations instead use a disc [26] or array [18] of pinholes similar to those used in some forms of confocal microscopy, or even in one imaginative but not terribly successful case a plastic slide of coloured stripes [23].

In addition to their particular drawbacks, there is a more general problem with the use of fixed componentry, which is that it trammels what is possible to even *consider* doing with the system. Standard diffraction gratings, for example, are inherently one-dimensional. Pinholes are of fixed size. Although we have only briefly touched on other possibilities here, it is very likely that removing these preconceptions will allow scientists to come up with whole new ways of generating the sample frames that may provide significant improvements to the process. There thus appears to us a very strong argument for replacing the prevalent opto-mechanical pattern generation techniques with a digital alternative.

The notion of using a *spatial light modulator* (SLM) for this purpose has been proposed before, and at least one successful implementation using a digital micromirror device (DMD) has been published [10]. The general-purpose structure suggested by Heintzmann *et al* [20], in particular, is very similar to our own. However, we are not aware of anyone to date having successfully implemented such a system using liquid crystal components.

5 Liquid crystal displays

5.1 Component structure

Liquid crystals are materials with a strongly directional molecular character that leads them to preferentially adopt orderly arrangements over some portion of their liquid phase. This is in contrast to 'normal' materials, whose liquid phase is isotropic.

The forms of order exhibited vary between different liquid crystal *mesophases*. Of particular importance are *nematic* liquid crystals, which show orientational ordering but not positional, and *smectic* liquid crystals, which show both [6].

The preferred orientation in a liquid crystal is known as the *director*. Molecules will tend to align that way because it is energetically favourable to do so; other orientations incur an elastic penalty. This can be exploited mechanically at the boundaries of a liquid crystal by making the contact surface force a particular orientation; that orientation then propagates within. The director can also be reoriented by applying an electric field, which encourages the molecules' dipole moments to line up in the field direction. With both mechanical and electrical forces in action, the director will depend on their relative strengths. This allows liquid crystal components to be produced that switch between known orientations according to an applied voltage.

Because a liquid crystal is anisotropic, it has different permittivities—and thus different refractive indices—in different directions; it is *birefringent*. This means it can, depending on its orientation, change the polarisation of light passing through it. Combined with the ability to control the orientation electrically, this allows liquid crystals to be used for display components (LCDs).

The basic structure is shown in Figure 21: a layer of liquid crystal is sandwiched between arrays of electrodes, which define the controllable pixels. Light is linearly polarised before it passes into the display, and its polarisation is either rotated or not according the the state of the electric field applied across each pixel. On the other side, a second polariser blocks the unrotated light, leaving only the light from the 'off' pixels.[6]

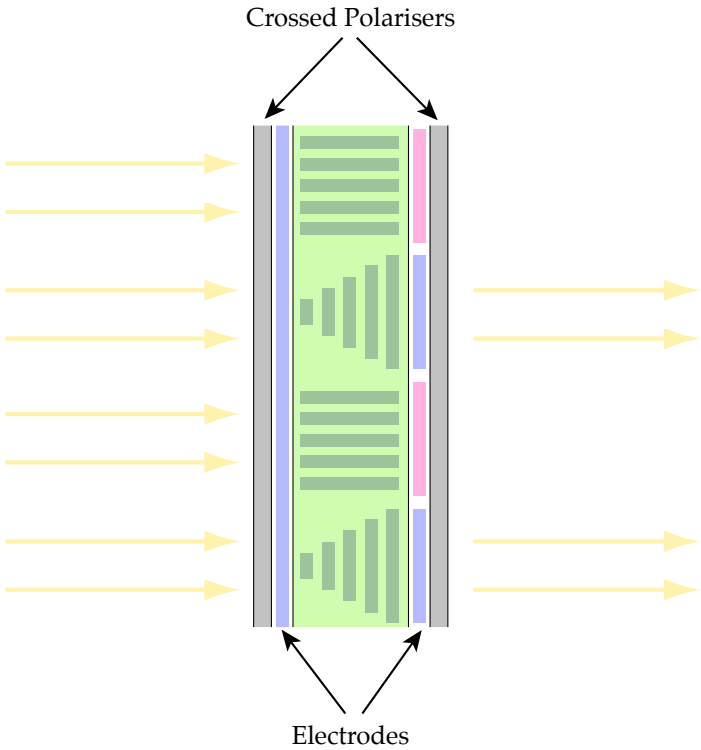

Figure 21: Structure of a twisted-nematic transmission LCD. Light only passes if its polarisation is rotated by the liquid crystal.

[6]The polarisers can equally be arranged to pass light from the 'on' pixels.

For displays with many pixels, additional electronics are needed at each one to manage its state. While this is practical in transmission mode for large devices such as laptop screens, it is less so for displays at smaller scales. The best current technology for manufacturing small, high precision LCDs suitable for use as SLMs in applications such as video projection and our own proposed SI design is *liquid crystal on silicon* (LCOS). This uses the liquid crystal in reflection rather than transmission mode, placing the control circuitry on an opaque, highly-polished silicon backplane. Because the incident and modulated light are both on the same side, a beam-splitter is required, complicating the illumination pathway somewhat.

The majority of LCDs use nematic liquid crystals, typically in a *twisted* configuration. In these displays the director varies more or less continuously between the 'off' and 'on' states with the applied electric field, so the pixels can be set to intermediate greyscale values. It is also possible to produce displays using *ferroelectric* liquid crystals, based on a chiral form of smectic mesophase [7]. Ferroelectric LCDs are *bistable*—an electric field is required to change their state, but not to maintain it either way. They have the advantage of considerably faster transition times—on the order of microseconds as opposed to milliseconds for twisted nematic—and hence can operate at higher frame rates, but they do not support intermediate states. They also usually have to be driven in a DC-balanced fashion, meaning that the average voltage imposed over timescales of 100ms-1s should be zero. This is achieved by rapidly alternating positive and negative frames and blanking the illumination during the latter phase [22], which may be problematic in the context of fluorescence illumination.

5.2 Spatial light modulation for structured illumination

We now consider the implications of using an LCOS SLM to generate the excitation patterns for SI microscopy.

Although our goal is the modulation of intensity, LCDs achieve this via modulating polarisation. As a result, we are constrained to working with linearly polarised light. Since the light sources for incoherent fluorescence microscopy are normally unpolarised, the initial polarisation will eliminate roughly half of the illumination before it even gets to the LCD. Consequently, even in the bright regions of the excitation pattern, the sample will be illuminated more dimly than with direct widefield illumination. For a given source, therefore, the fluorescence will be weaker and either the exposure times will need to be increased, or else the SNR will be lower. The impact of this will, of course, depend on the degree of fluorescence in the sample—for brighter samples we may already need to introduce filters into the illumination pathway to reduce saturation and photobleaching, in which case the reduced intensity will not be a problem.

The discarded light must go somewhere, ultimately to be absorbed by the equipment or its environs, causing localised heating. This will need to be managed carefully to avoid compromising the precision of the pattern projection, but should be considered in the context of widefield epi-illumination, in which a fair amount of the excitation light is also absorbed along the way without major ill-effects.

In addition, there is the possibility that the fluorophore distribution in the sample may be anisotropically responsive to polarised excitation, in which case the emissions produced will be biased by the polariser direction. This is quite unlikely in most biological applications, where the distribution should be approximately isotropic and the number of participating fluorophores large. While there may be specialised cases where it becomes an issue, we have not been able to identify any important ones. The most likely context in which this could arise is with membrane proteins, which have at least some orientational anisotropy on account of

being embedded in the membrane. However, these will typically still have rotational freedom and hence at least quasi-isotropic orientation of fluorophore dipole moments.

The principal benefit of using an SLM over existing optomechanical implementations is that the details of the modulation can be varied at will. The bounds on this versatility are those of the control circuitry and of the individual addressable elements of the device, its pixels.

Although there is no absolute requirement that this be the case, LCD pixels are almost universally rectangular themselves and arranged in a rectangular grid; any other configuration would need to be custom built at significant expense. We will therefore assume such a configuration here, but noting that the assumption could be relaxed were there a strong enough argument for doing so.

The area of each pixel is defined by the electrodes that generate the controlling field. Clearly these cannot be continuous; each must be insulated from its neighbours, so there are gaps between the pixels. The liquid crystal, however, *is* continuous across the gaps. Where neighbouring pixels are driven to the same state, that state will also predominate in the gap. Where they are in different states, there may be some transitional region in between. The gaps are typically at least an order of magnitude smaller than the pixels.

Since our goal is to project an excitation pattern with components on the same order of detail as the optical diffraction limit, we are interested in displays whose pixels are as small as possible. Details of some current commercially-available LCOS components are listed in Table 2. We observe that the pixel pitches are of the order 10μm and will take this as a convenient scale estimate. It is likely that fabrication technology will improve over the next few years; in any case, it will not get worse, so this is a reasonable conservative value.

Device	**Pixel Pitch (μm)**	**Gap Size* (μm)**	**Number of pixels**	**Contrast ratio**	**Switch time (ms)**	**Refresh rate (Hz)****
4DD SXGA-R2D	13.62	0.5	1280×1024	n/a	0.04	60
4DD SXGA-R3	13.62	0.5	1280×1024	n/a	0.04	85
HOLOEYE BR768HC	12	0.4	1280×768	1800:1	7-11	120
HOLOEYE BR1080HC	8.1	0.4	1920×1200	2000:1	7-11	120
HOLOEYE Z86D-3	12	0.4	800×600	100:1	n/a	120

* Size calculated from published fill ratio.

** When driven as full-colour video.

Table 2: Manufacturer's specifications for some current LCOS components [9, 21].

The image of the LCD will be projected onto the specimen via the microscope objective, so it will be demagnified in proportion to the objective's power. There is some scope for controlling this relation, as will be discussed in §7, but for now let us assume that the scaling down is the same as the nominal magnification. On this basis, Table 3 compares the approximate pixel pitch at the specimen with the Rayleigh criterion resolution limit (equation 1) for several typical objectives with emission light at 520nm. Also included is the optimum grating wavelength for optical sectioning described §4.1.

The demagnified pitch is at least of the right order of magnitude, and indeed for magnifications of $20\times$ and greater the individual image pixels will not be resolvable. (By the same token, the much smaller inter-pixel gaps will be far beyond the resolution limit and their contribution can be safely ignored.) We discuss how this relates to our patterning requirements in §5.3, below.

Note that the area of the specimen that can be illuminated will depend on the size of the entire SLM. Since demagnification for illumination and magnification for imaging are both performed

Objective magnification	NA	Resolution limit (nm)	Sectioning optimum (nm)	Pixel pitch (nm)
10×	0.35	900	1400	1000
20×	0.55	570	940	500
40×	0.75	420	690	250
60×	0.95	330	550	166
100× (oil)	1.40	225	370	100

Table 3: Demagnified pixel pitches compared. Resolution limit is for 520nm emissions.

by the same objective, the portion of the field of view covered is independent of the objective and instead determined by the *field number* of the eyepiece. This is typically around 16-20mm although somewhat larger fields are possible [27].

The SLM active dimensions can be readily calculated from Table 2; considering only the shorter side, we find they range from 7.2mm for the low-end Z86D-3 to 14mm for the 4DD components. For a demagnification ratio of 1, then, our excitation patterns—and the corresponding reconstructed images—will not cover the whole field of view.

In addition to the spatial resolution of the SLM, we are also interested in its time resolution, since one of our main goals is to increase the speeds at which it is possible to obtain SI images compared to existing implementations. Referring again to Table 2, we see that switching times vary markedly between displays; in particular, the ferroelectric 4DD components are orders of magnitude faster than the Holoeye twisted nematics. The former (or their successors) may therefore offer an advantage in the long term if high frame rates become necessary, despite the awkward DC balance requirement mentioned above.

However, as we'll see in §6.2, the achievable frame rate is more likely to be constrained by the exposure time for image capture than the speed at which the LCD can be switched. And while we have no present use for colour patterns in SI, the ability to use standard hardware and software interfaces to drive the SLM will likely make the supported video refresh rates a more relevant speed metric in the short to medium term.

Taking the slowest advertised rate as our basis, we should be able to project up to 60 excitation patterns each second. The corresponding frame rate for reconstructed images will obviously depend on the illumination strategy and method of reconstruction. For depth sectioning, with its requirement of just three frames per recovered image, this equates to a respectable 20 frames per second, which would represent a fourfold increase over the best speed of the Zeiss Apotome (and consequent ability to view biological entities moving four times faster). For lateral resolution improvements requiring six to nine captures the frame rate will obviously be correspondingly less.

5.3 Patterns as pixels

The sorts of excitation patterns discussed in §4.3 were principally either plain sine gratings or simple combinations of a small number of frequency components. Indeed, our sampling efficiency depends on this, since a pattern with more frequency peaks will require more frames to disentangle.[7] When generating our illumination pattern with an LCD, however, we are not

[7] There is an exception: the max-min algorithm is readily applicable to complex patterns provided they sample both dark and light thoroughly. It would work very well with perfect square wave illumination, if we were able to project such a pattern.

directly generating such smoothly-varying patterns. Instead, we are constrained to work with the pixel grid.

This turns out not to be a major problem for the patterns themselves, because we are aiming to work with spatial frequencies reasonably close the diffraction limit. By definition, the higher diffraction orders in these patterns will be outside the region of support of the OTF, so only the fundamental frequency of the pattern will be imaged.

However, we must still negotiate the matter of shifting the pattern phase, and here the quantisation is significant: we can only move the pattern by a whole number of pixels. Thus, for a pattern with period n pixels, the only possible phase shifts are multiples of $\frac{2\pi}{n}$. In order to produce the $\frac{1}{3}$-cycle shifts presupposed by the Neil algorithm, for example, our pattern period must be a multiple of 3 pixels.

Referring once again to the values in Table 3, we see our minimum period for this kind of illumination—using an illumination pattern of one on pixel to two off—ranges from 3μm for the $10\times$ objective to 300nm for the $100\times$. This should still be sufficient for some sectioning benefit, but we do not achieve the optimal grating pitch for any of these example objectives. We may be able to optimise this for a particular objective by adjusting the demagnification (see §7), but there will always be some degree of compromise with so few pixels to play with.

Similar considerations apply to two dimensional patterns: we can only adjust in pixel increments in either direction. If we want to have an even set of phase shifts, we will need sufficient pixels both ways. There is one minor wrinkle, however, which is that the dominant spatial frequencies of the most pixel-friendly two dimensional pattern, a checkerboard, occur along its *diagonals*. Where these are the only frequencies to make it through the OTF, the equivalent pitch of the constituent sine gratings is reduced by a factor of $\sqrt{2}$. Whether this is beneficial will largely depend on how the available pixel pitch compares to the ideal already.

6 Computational requirements

Since the images from SI must be algorithmically reconstructed rather than being directly available to the eye, much of the business of any SI system will be computational. The key computer responsibilities are outlined below. We note that there is more or less infinite scope for finessing the software side to provide the user with ever more sophisticated tools for analysis and control. We do not propose to concern ourselves with such details here beyond a few very general observations.

6.1 Pattern generation

One of the principal computer functions will be to drive the SLM, generating the appropriate sequences of excitation patterns. However, this is expected to be mostly administrative. While there may be some engineering difficulties in actually interfacing to the hardware, the patterns themselves—at least, those which have been considered to date—will be computationally trivial to generate. We do not expect to encounter any problems of complexity or performance in this regard.

The main administrative concern would be managing the synchronisation of the pattern generation with the potentially more problematic issue of acquiring data from the CCD.

6.2 Data acquisition

In order to reconstruct our final images, we need to acquire a sequence of frames from a CCD camera. The specifications of the latter have not been examined in any detail in this project, but we make the following observations.

The captured image resolution needs to be such that it samples the transferred emissions at a high enough rate for the variations with the excitation pattern to be identified. Since the pattern is bandlimited, the Nyquist-Shannon sampling theorem applies and the sampling rate should be at least twice the maximum spatial frequency in the excitation pattern, k (which we expect will be of roughly the same order as the cutoff frequency k_0).

Assuming that the pattern has been generated by an LCOS device similar to those discussed in §5.2, with an excitation strategy using $n \geq 3$ pixels per period as indicated by §4.3, then to accurately capture the patterned emissions the CCD must have at least $\frac{2}{n}$ as many pixels measuring that portion of the field of view as the SLM has illuminating it. Taking $n = 3$ and a typical SLM pixel dimension of 1280, this corresponds to ~ 850 CCD pixels, so we would likely want at least a 1024×1024 pixel resolution on the CCD camera.

Depending on the number of bits recorded per pixel, this corresponds to a significant quantity of data per frame: 1MB for a standard 8 bit greyscale, 1.5MB for the more sensitive 12 bit. This data must be retained on the CCD until it can be transferred to the computer, delaying the start of exposure for the next frame. Thus both the communications latency and the exposure time limit the achievable frame acquisition rate.

In some CCDs it is possible to overlap these two processes and thus be limited only by whichever is slower. This is done by using some portion of the onboard pixel array memory as a storage buffer. Captured values are transferred to neighbouring pixels, which is a fast parallel operation. The original pixels can then start capturing the next frame while the slower transfer of the frame data off-chip takes place from the buffer pixels. The cost of such a strategy, obviously, is the sacrifice of capture resolution, because only a fraction of the available pixels are used.

Two versions of this scheme are found. In *inter-line transfer* CCDs, alternate lines of pixels are used as buffers, which can lead to line spacing artefacts that would be undesirable for SI. In *frame transfer* CCDs, separate regions are used for capture and buffering. The latter arrangement is clearly preferable for our purposes, assuming the capture region still includes enough pixels for successful sampling.

Although the bandwidth requirements are a concern, commercial CCD cameras are available capable of very high transfer rates. The principal limiting factor is therefore likely to be the integration time for each frame. This is much harder to quantify, as it will depend intimately on the nature of the specimen and the fluorophore concentrations present.

On the basis of some experimental images captured by colleagues in the Pharmacology lab, we have estimated that exposures of 20ms produce images with sufficient detail for further analysis, while those at 10ms do not. From this we would expect to capture individual frames at no more than 50 Hz. These results are little more than anecdotal, and there may be plenty of scope for variation, but they at least provide us with one initial data point.

6.3 Image processing

While interfacing to the various other components may be a significant engineering task, the major computational challenge is that of reconstructing complete images from their contribu-

tory frames. We anticipate that this will be implemented wholly in software on CPU to begin with, which is likely to limit real-time processing to (at best) the simpler algorithms.

This is not ideal, but we note that provided the captured frame data are stored for subsequent analysis we are not limited to online processing. Similarly, an important advantage of the sort of flexible system we are proposing is that there is no need to restrict ourselves to a single processing technique, as is done in commercial implementations such as the Apotome. There is at least some scope for switching techniques as appropriate to the experimental context.

Taking as our basis the admittedly-dubious estimates from the previous section, we can get some rough idea of the computational difficulty of the reconstruction task. Each data frame consists of 2^{20} pixels; many of these may be of no interest, but it is safer to assume we don't have any *a priori* way of discarding them. For simple 3 frame sectioning, we need to combine each pixel with its counterparts from the two other frames.

While we have expressed both the Neil and max-min algorithms algebraically, it may be more efficient to implement them via a look-up table if the capture bit depth is not too large—for 3 8 bit values the table would require 2^{24} entries or about 16 MB for a single byte result. The computational cost of the algorithms would then be small—if we say a nice round 16 CPU operations per final pixel, this requires 2^{24} ops per frame or about $2^{28} \approx 3 \times 10^8$ per second for 50Hz captures. This is well within the theoretical capabilities of a modern CPU, although in practice the process is likely to be heavily I/O-bound. If memory access times become prohibitive, a vanilla software implementation even of these simpler methods may not be able to run in real time. However, both algorithms are perfectly suited to implementation on massively parallel graphics hardware and doing so should not be very much more difficult.

The more complex linear algebraic method, at least in its general form, is vastly more computationally expensive, requiring $n + 1$ Fourier transforms for n frames, as well as the inversion of a very large number of $n \times n$ equation systems. In this case it is doubtful that real time performance could be achieved with anything less than dedicated hardware. However, offline processing could certainly be used on stored frame sequences.

7 Proposed instrument design

As already discussed, the modern optical microscope is a complex and high-precision piece of equipment and we do not intend to reinvent it. The benefits possible through SI are adjunct to the qualities of the microscope itself and the technique should be deployable as an add-on option rather than a radical rebuild. We here describe some of the requirements for such an add-on and suggest a possible design.

7.1 Practical patterned illumination test

While the idealised description of the structure of an infinity tube microscope given in §3.3 is true as far as it goes, any actual microscope will also have a host of specific details and features that must be taken into account when attempting to produce illumination patterning. To investigate the feasibility of imposing such patterning in reality, we added a simple projection set-up to one of the Pharmacology Department microscopes. Some interesting quirks were encountered in the process, which inform the design suggestions in the next section.

Since the experiment was intended solely to test the pattern projection, we used a static pattern, cast in this case by a commercial NBS 1963A resolution target. In order for the objective to

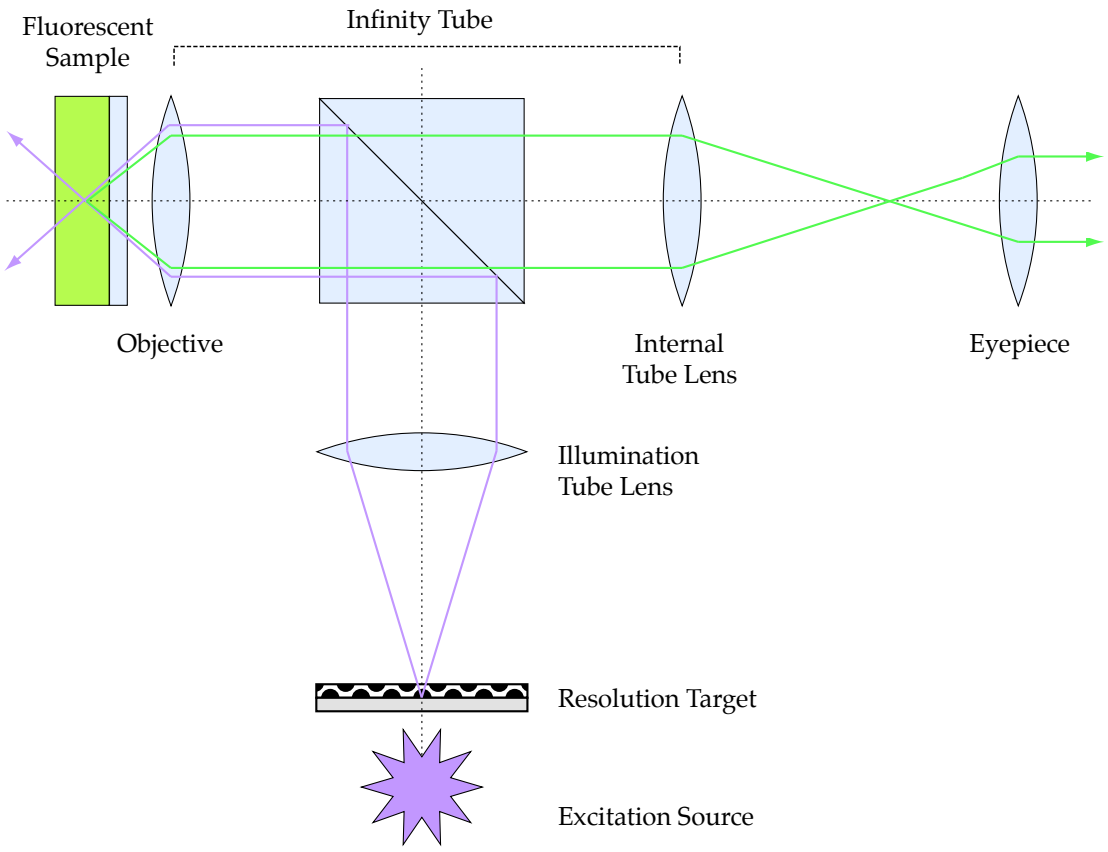

Figure 22: Experimental arrangement for excitation with a fixed target pattern.

focus the image of this target at its focal plane, the light from the target had to be parallel in the infinity tube region, collimated by a tube lens similar to the one delivering the objective's image back to the eyepiece (Figure 22).

The microscope used, a Leica DM IL, is a basic inverted infinity tube model with a removable mercury arc housing at the rear to provide fluorescence excitation. In order to insert the resolution target and tube lens into this illumination pathway, the housing was moved back and an extension added so that the lamp could still be used safely. However, the process of setting up the target with the arc lamp was extremely unwieldy and most of the experiments were done instead with a safer halogen source and no housing, which proved quite adequate to excite the specimens used.

The target and tube lens were mounted on a metal base, with a rack and pinion mechanism allowing the target position to be adjusted (Figure 23). Initial tests used a lens of focal length 200mm, the same as the microscope's internal tube lens, for demagnification equal to the nominal magnification of the objective. Later tests were also done with a 50mm lens for one quarter the demagnification.

The specimen onto which the pattern was projected was a thin layer of fluorophore solution sandwiched between a glass slide and cover slip, or in some later tests a thick volume of the same solution contained in a dish. In both cases the purpose was to examine the light pattern rather than the details of the sample.

The first attempts to produce a focussed image were unsuccessful due to the presence of two

Figure 23: Projection apparatus for the imaging the NBS 1963A resolution target.

optical elements in the illumination pathway: a positive lens at the external illumination port and a ground glass diffuser glued into the interior body of the microscope just before the dichroic filter block. The former was part of the instrument's back panel and easily removed; under other circumstances it might instead have been pressed into service in place of the tube lens. The diffuser was a mystery to all, including visitors from the microscope companies, its presence entirely at odds with the orthodoxy of Köhler illumination, rendering any kind of illumination patterning impossible. It was also very firmly attached and could only be removed by brute force.[8]

With the obstacles removed, an image of the target could indeed be projected on the sample, as seen in Figure 24. It's noticeable that the image formed with the higher magnification lens is in considerably poorer focus. Much of the problem is down to inadequacies of alignment and calibration in the test rig, which is to be expected given that the equipment was fairly crude and aligned by hand and eye. However, at least some of the fault may also be due to a more fundamental limitation: the resolving power for image projection depends not only the objective, but also on the tube lens. We will return to this issue in the next section.

Figure 24: Fluorescence images of the projected resolution target at 10× (left) and 40× (right) demagnification.

Despite some success in projecting our illumination pattern, the experiment revealed a more puzzling phenomenon that we have been unable to satisfactorily explain. Consider the optical setup as depicted in Figure 25. The objective will focus parallel light at its focal plane.

[8] This project will thus, if nothing else, be the envy of the many experimental biologists over the years who have wished to whack their microscopes with a hammer.

Convergent light will focus closer than that plane, divergent light further away. If the target is positioned correctly with respect to the tube lens, its light will be parallel and it should remain in focus regardless of the position of the objective; if, on the other hand, it is incorrectly positioned, its light will be either convergent or divergent and it will always be out of focus, projecting either to one side of the objective's focus or the other.

What we in fact find, as shown in Figure 26, is that the projected image moves in and out of focus as the objective is moved—and that the in focus image appears to *pass through* the objective's focal plane. A number of hypotheses have been suggested to explain this, but most have been tested and failed. The current best remaining guess is that we are being misled by treating the objective as a single lens when it is in fact a compound of many optical elements. With slightly divergent, convergent or asymmetrical light it may behave unexpectedly as paths cross internal element boundaries. However, we note that microscope lens manufacturers put a good deal of effort into making objectives act like a single indivisible unit, and have to hope that a better explanation can be found.

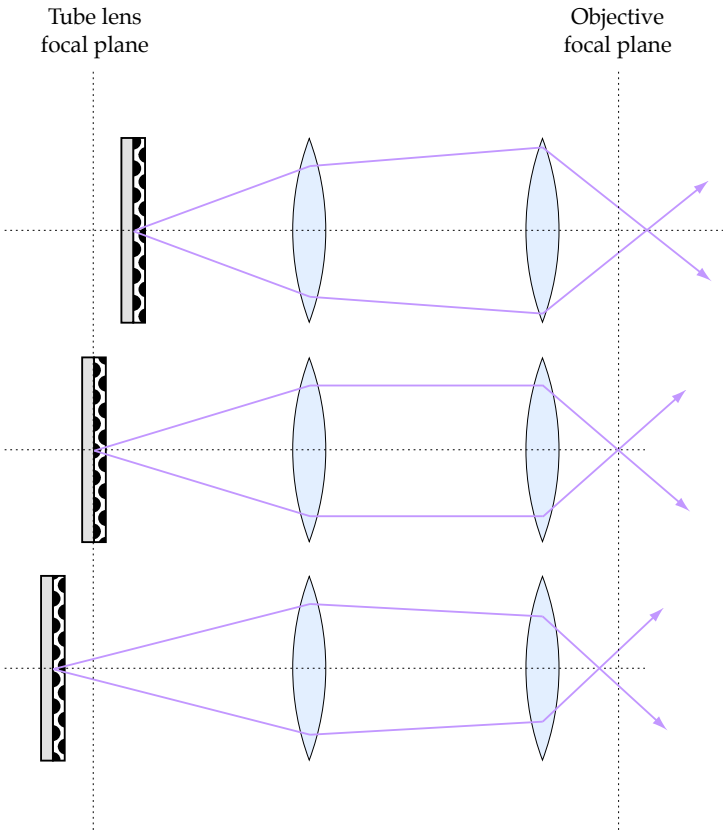

Figure 25: If the light from the target is not parallel when it reaches the objective, it will be focussed away from the objective's focal plane. Moving the objective may bring the image closer to the focal plane, but it can never reach it. Certainly, the image should not be able to cross from one side to the other.

Although this phenomenon is unfortunate, it doesn't present an insurmountable obstacle to implementing SI, since the pattern can focussed to its optimum by moving the objective, and the focal plane then shifted through the specimen by moving the stage, leaving the objective where it is. This is less convenient for those laboratory microscopes that are designed to focus by moving the objective (like this Leica), but is a very common approach in microscopy—indeed, the same technique is used for SI sectioning by the Zeiss Apotome.

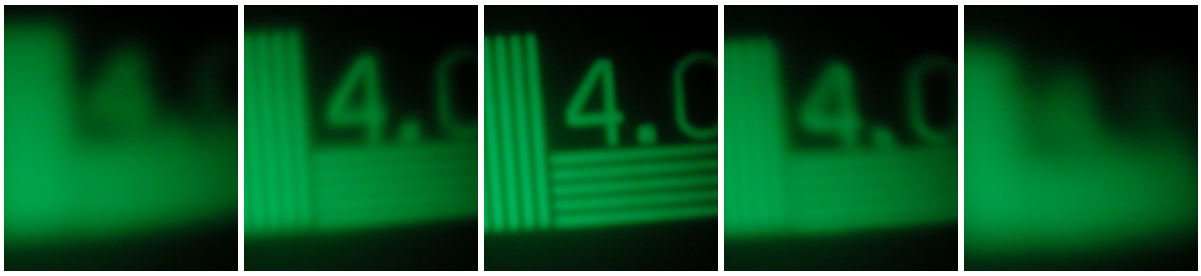

Figure 26: The projected pattern goes in and out of focus as the objective moves.

7.2 Physical structure

The basic arrangement of elements needed for SI microscopy using an LCOS SLM is shown in Figure 27. Similar arrangements have been presented in the literature, notably by Heintzmann *et al* [20] and Fukano & Miyawaki [10].

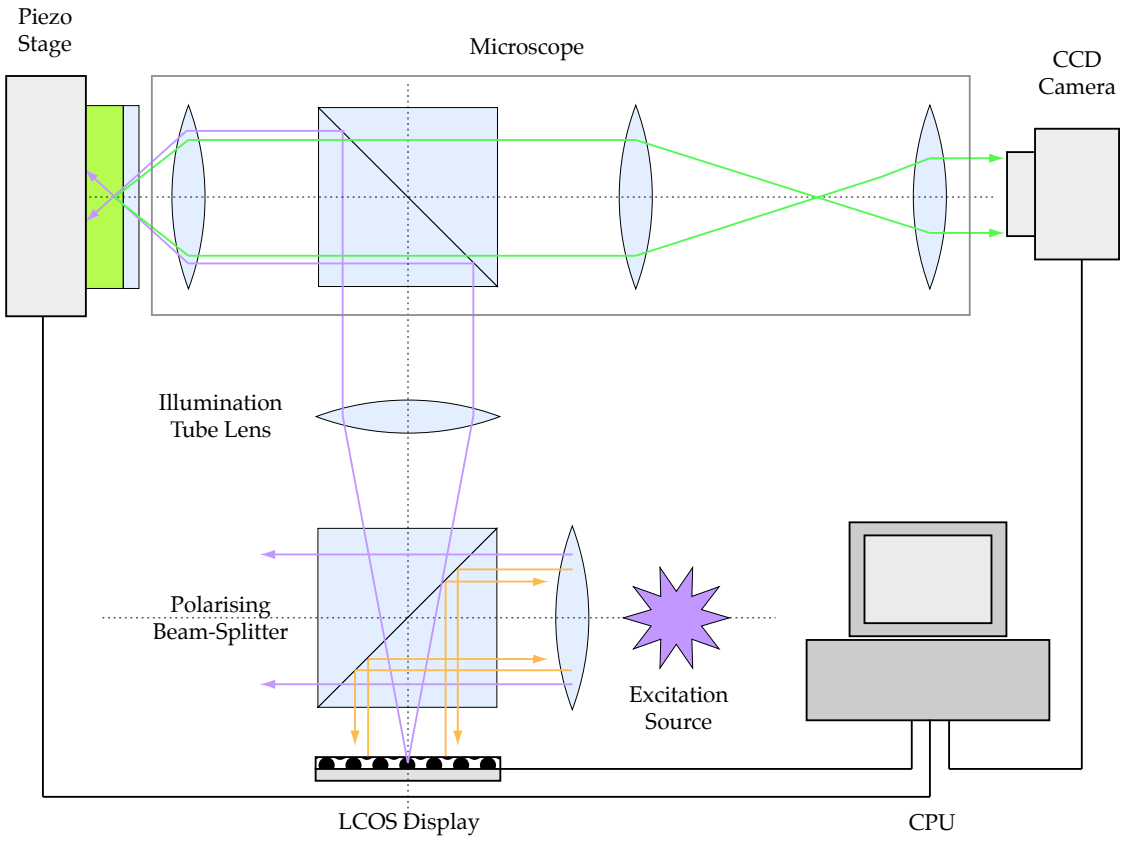

Figure 27: General structure of an LCOS SI microscope.

For practicality we assume that the main body of the microscope to which SI functionality is to be added will not be substantively altered. The major elements under our control are therefore the SLM itself and the illumination tube lens, along with the CCD camera and the computer hardware and software driving the SLM and processing the images.

In discussions so far we have taken as a given that the illumination tube lens will be matched to the microscope's own, resulting in demagnification equal to the nominal objective mag-

nification. However, this is not a requirement and there may be good reasons for choosing otherwise.

Tube lens focal lengths for infinity optics microscopes vary from manufacturer to manufacturer: Nikon and Leica use a tube lens of 200mm, Olympus 180mm, Zeiss 165mm [27]. Denoting this f_t and the objective nominal magnification M_o, the effective objective focal length f_o is simply

$$f_o = \frac{f_t}{M_o} \tag{20}$$

The effective demagnification of the excitation pattern, M_i, for an illumination tube lens of length f_i is then

$$M_i = \frac{f_i}{f_o} = \frac{f_i}{f_t} M_o \tag{21}$$

Evidently, we can choose any demagnification level we like simply by varying f_i. However, there is a cost: to increase the demagnification we must increase the distance between the SLM and the tube lens. In doing so, we lose more and more of its light and, crucially, of its diffraction orders, resulting in a dimmer, more poorly-resolved excitation pattern. While we can improve matters to some extent by using a larger-diameter lens, this will entail adding other optical elements to condense the captured light since the eventual collimated beam needs to pass through the aperture of the infinity tube. Moreover, the lens size is unlikely to come anywhere near the 165-200mm scale of f_t; nor is a high refractive index medium between the SLM and the lens practical. So the NA will always be much lower than the objective.

There is thus a trade-off between the amount of demagnification and the quality of the image produced—or, at least, the complexity and precision of the optics required to produce a satisfactory image. It should be clear from the discussions of §§5.2-5.3 that a significant level of demagnification is necessary if we are to achieve a meaningful benefit from the use of SI, given the SLM pixel sizes available. Equally clearly, this amount should be kept to the bare minimum needed and no more.

In moving from the abstractions of our schematic to a physical embodiment, one thing that our pattern projection experiment demonstrates is the awkwardness of integrating with the existing structures of the microscope's illumination pathway. It would be much more convenient if we could effectively add an entirely new pathway for our own purposes without having to meddle with those already present. Ideally, we would be able simply to plug our add-on into the microscope—indeed, into any microscope we choose—with a minimum of fuss or bother.

While variations between manufacturers and even microscope models makes the goal of universality unrealistic, it turns out that there *is* a route that we can use for fuss-free modular integration: the one used to hold the swappable filter blocks (Figure 28).

In normal operation, the filter block is angled towards the internal illumination source, and the sliding metal housing that contains it has its outward-facing surfaces sealed to avoid leakage of the excitation light. If the filter block were instead rotated by 90°, it would work with illumination coming from the side. A unit of the same cross-section as the housing but with the necessary arrangement of SLM and tube lens built inside it would provide just the modularity desired. It would also have the additional benefit of being positioned at the side rather than rear of the microscope; the latter is often inaccessible and limited in space.

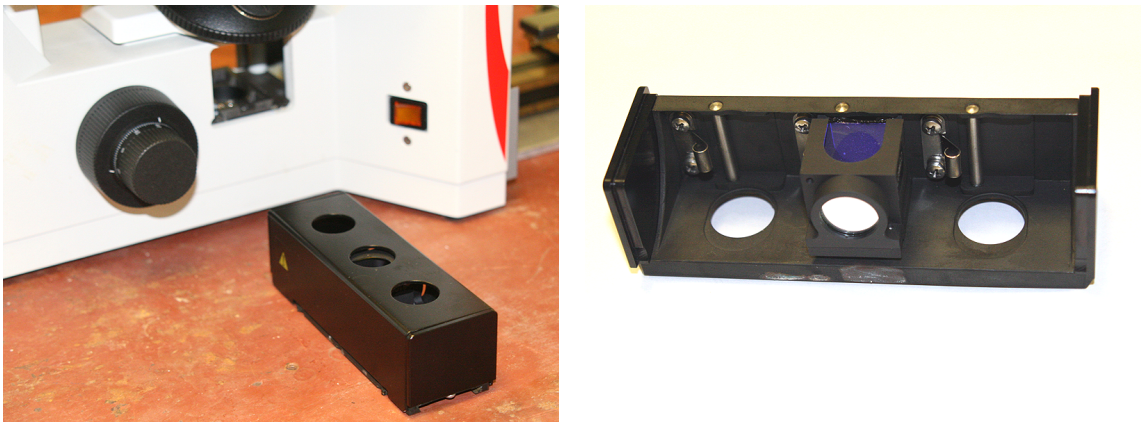

Figure 28: The filter block housing provides a potential port for a modular patterned illumination source.

Because of the focus problem described in the previous section, we require a piezoelectric stage or similar mechanism for moving the specimen relative to the objective's focal plane. Such equipment is used widely in other microscopy applications and is readily available from commercial sources, so we do not address it further here.

We similarly suggest that the computer hardware and CCD camera used will be off-the-shelf commercial units, although the choice of the latter must take into consideration the particular concerns outlined in §6.2. In the worst case, some custom development may be required here.

As discussed in §6.3, image post-processing will initially be performed purely in software. While this may be feasible in real time for the simpler reconstruction algorithms, we believe it will not permit full reconstruction by the linear algebraic method. In the short term, offline processing will be needed to reap those benefits and to confirm the efficacy of the method. Subsequent reimplementation using custom hardware may be able to support real time linear algebraic recovery if it proves desirable to do so.

7.3 Discussion

Our investigation into the possibility of producing an SI microscope using an LCD for pattern generation suggests that such a microscope is probably feasible, and our proposed design proceeds on that basis. A number of difficulties have already been identified. There are also several areas that are lacking a great deal of detail and it is possible—even likely—that there will be further obstacles associated with these. Even so, we consider this a reasonable position from which to begin to assess whether such a microscope is likely to provide tangible benefits for biological research.

Revisiting the resolution constraints discussed in §2, we can make the following observations.

Using components of the kind described above, appreciable benefits should be achievable in terms of depth sectioning at speeds significantly in excess of those for existing implementations. This would represent a distinct advance in experimental utility and would, for example, support an application such as that described in §2.3 for imaging different cell compartments simultaneously.

The prospects for lateral resolution are marginally less rosy. Although it should still be possible to achieve some improvement, the requirement for larger numbers of contributory frames means that time resolution will remain relatively low, even putting aside the computational difficulties of real-time reconstruction. This is less disappointing when viewed in the light of other techniques available for the same purpose, since none are without their own downsides. We also note in particular that the edge enhancement effects of the lower cost algorithms may be advantageous for applications where it is more important to be able to separate biological entities than to image them faithfully, which is often the case when using fluorescent probes.

It must be borne in mind that one of the key benefits of using an LCD, aside from potentially increased speed, is its versatility. We believe this should be a boon in itself, allowing the experimenter to switch between different modes or recovery methods at will. The ability to select between several different section thicknesses, for example, might be very helpful when attempting to establish the relative positions of different fluorescent objects within a cell.

However, there is also an evident cost associated with this versatility: in attempting to cater for a variety of uses, the system will not necessarily be optimised for any of them. We saw this in §5.3, with the quantisation of our grating patterns to the pixel grid. A custom glass diffraction grating could easily be made to be the perfect pitch for one particular objective. We believe that this possible disadvantage is substantially outweighed by our system's flexibility. In the case where one particular configuration is of especial importance it should still be possible—if perhaps not convenient—to optimise the system for that purpose.

8 Future work

While this project has examined some of the issues involved in producing an SI microscope using an LCOS SLM, we have in a sense only begun to scratch the surface and there is no end of further work to be done. This falls broadly into two categories.

In the *engineering* category, our initial design needs to be refined and evaluated to see whether it is, as we believe, a worthwhile proposition. Ultimately, it would need to be specified, built and applied to some substantive biological problems—only then will we know whether it can make a genuine contribution to research.

Such an endeavour will be complicated, fraught with difficulty and, despite what has already been learned, still somewhat speculative. But if successful it will provide an excellent foundation for further refinements, for example extending it to support the non-linear saturated model for further lateral resolution improvements. These in turn should allow biologists to explore the processes of life in ever-greater detail.

The other category of future work is the purely *theoretical*. One obvious problem that needs to be properly elucidated is that of the edge enhancement effects described in §4.5. This problem is on a very different scale to that of engineering elaborate experimental apparatus, but it is in its own way just as fascinating and could also prove to have practical applications.

Beyond that, the whole question of how the excitation patterns might be more optimally applied once freed from the physical constraints of fixed gratings is ripe for exploration. One question that has not even been considered is whether there is any reason to stick to phase shifts of a single pattern. Perhaps there may be better strategies in which several completely different patterns are used in combination. There has been little incentive to explore such questions while we have depended on fixed components for pattern generation. The prospect of an

illumination system that is not so constrained gives these problems a new immediacy.

9 Conclusions

Structured illumination offers clear benefits to users of optical fluorescence microscopy. Although there are corresponding disadvantages of increased complexity and lack of transparency, the burden compares favourably with that for other widely-accepted techniques such as scanning laser confocal microscopy. A system that provides this capability in a more flexible manner than is presently available would be of great value.

We believe that LCOS display technology provides a plausible way to implement such a system and that doing so merits further attention. The resolutions available with present generation devices are sufficient to obtain some benefit, especially in the matter of depth sectioning, although there are also significant limitations that leave scope for improvement as the technology advances.

Building an LCOS SI microscope will not be a trivial undertaking. In particular, the levels of demagnification currently required will demand rather high levels of precision in the construction and configuration of the illumination system. The apparent dependence on a moving specimen stage is also inconvenient, though that is consistent with many of the ways in which we envisage such a system being used.

Nevertheless, the ability to project illumination patterns freely will offer immediate advantages to experimentalists, and by opening up the possibility of previously unexplored excitation strategies may provide much greater benefits in the long term.

A Selected source code

The following are the main MATLAB functions used to generate the simulation images of §4.6. A package containing all the M-files may also be downloaded from the project website `http://www.ucl.ac.uk/~ucbpmbc/project.html`. All code was run under MATLAB version 7.4 (R2007a) with the Image Processing Toolbox and Signal Processing Toolbox.

```
function a = overlap ( d )
%OVERLAP gives overlap area of two unit r circles, centres separated by d
% A = OVERLAP ( D )

a = real((2 * acos(d/2)) - (sqrt(1 - ((d/2) .* (d/2)) ) .* d));
```

```
function otf = circotf(d)
%CIRCOTF generates a centred, idealised OTF for a circular pupil
% OTF = CIRCOTF ( DIAMETER )

[f1,f2] = freqspace(d, 'meshgrid');

otf = overlap(sqrt(f1.^2 + f2.^2) * 2)/pi;
```

```
function [img2, ft] = otfblur(img1, cut)
%OTFBLUR performs an idealised OTF blur
% [IMG2, FT] = OTFBLUR ( IMG1, PROPORTION )

if nargin < 1, img1 = target(); end
[h w] = size(img1);
if nargin < 2, cut = 0.5; end

% pad to doubled square
pad = 2 * max([h w]);
t = fft2(img1, pad, pad);

% diameter of OTF (doubled because of padding)
d = floor(max([h w]) * cut * 2);
otf = circotf(d);

% centre in black field
f = zeros(max([pad d]));
off = floor((pad-d)/2);
if off < 1, off = 1; end

f( off:(off+d-1), off:(off+d-1) ) = otf;
[h1 w1] = size(f);
clip = floor((h1 - pad)/2);

if (clip > 0), f = f( clip:(clip+pad-1), clip:(clip+pad-1) ); end

ft = fftshift(f) .* t;

% invert transform and remove padding
img2 = real(ifft2(ft));
img2 = img2(1:h, 1:w);
```

```
function [p] = combine(i1, i2, i3)
%COMBINE diffs three images illuminated at different phases
% G = COMBINE ( IM1, IM2, IM3 )

p = sqrt( (i1 - i2) .^ 2 + (i2 - i3) .^ 2 + (i3 - i1) .^ 2 );
```

```
function img = excess(i1, varargin)
%EXCESS returns the max pixel less the min at each position
% IMG = EXCESS ( IM1, IM2, ... IN )
%
% Either pass the images as separate args, or else as a single-row cell
% array

if nargin == 1
    mx = i1{1,1};
    mn = mx;

    [h w] = size(i1);

    for ix = 2:w
        mx = max(mx, i1{1, ix});
        mn = min(mn, i1{1, ix});
    end
else
    mx = i1;
    mn = i1;

    for ix = 1:(nargin-1)
        mx = max(mx, varargin{ix});
        mn = min(mn, varargin{ix});
    end
end

img = mx - mn;
```

```
function [g] = vgrating(w, h, wavelength, ph)
%VGRATING creates a vertical sine grating in range [0,1]
% GRATING = VGRATING ( WIDTH, HEIGHT, WAVELENGTH, PHASE_OFFSET )

if nargin < 1, w = 100; end
if nargin < 2, h = w; end
if nargin < 3, wavelength = 10; end
if nargin < 4, ph = 0; end

step = 2 * pi / wavelength;
for x = 1 : w
    r(x) = (1 + sin(ph + (x - 1) * step))/2;
end
g = [];
for x = 1 : h
    g = [g;r];
end
```

```
function varargout = phases(w, h, wavelength)
%PHASES produces evenly-spaced phase-shifted vertical gratings
% [G1 G2 ... ] = PHASES ( W, H, WAVELENGTH )

if nargin < 1, w = 100; end
if nargin < 2, h = w; end
if nargin < 3, wavelength = 10; end

for i = 1 : nargout
    varargout(i) = {vgrating(w, h, wavelength, (i-1) * 2 * pi / nargout)};
end
```

```
function [f1 f2 f3 c0 c1] = sinsim3 ( obj, noise, period, cut1, cut2 )
%SINSIM3 simulates patterned excitation with a sin grating
% [F1 F2 F3 C0 C1] = SINSIM3 ( OBJ, NOISE, PERIOD, FRAC1, FRAC2 )
%

if nargin < 1, obj = target(); end
if nargin < 2, noise = 0.1; end
if nargin < 3, period = 5; end

[h w] = size(obj);

if nargin < 4, cut1 = 0.5; end
if nargin < 5, cut2 = cut1; end

if length(noise) > 1
    varnoise = noise(2);
    noise = noise(1);
else
    varnoise = 0;
end

[ch1 ch2 ch3] = phases(w, h, period);

f1 = obj .* otfblur(ch1, cut1);
f2 = obj .* otfblur(ch2, cut1);
f3 = obj .* otfblur(ch3, cut1);

n = blur(rand(h,w) * noise, 20);

f1 = otfblur(n + f1, cut2);
f2 = otfblur(n + f2, cut2);
f3 = otfblur(n + f3, cut2);

if varnoise > 0
    f1 = imnoise(f1, 'gaussian', 0, varnoise);
    f2 = imnoise(f2, 'gaussian', 0, varnoise);
    f3 = imnoise(f3, 'gaussian', 0, varnoise);
end

c0 = otfblur(n + obj, cut2); c = imnoise(f1, 'gaussian', 0, varnoise);
c0 = c0/max(max(c0));

c1 = combine(f1, f2, f3);
```

```
function [c0 c1 ff] = chksimnx ( obj, noise, tile, sub, cut1, cut2 )
%CHKSIMNX simulates supersampled excitation with a checkerboard pattern
% [C0 C1 C2 FF] = SQSIMNX( OBJ, NOISE, TILE, SAMP, FRAC1, FRAC2 )
%
% TILE is the size of single square in pixels
% SAMP is the number samples per cycle in one direction (ie, SAMP^2 samples
%      will be taken overall)
%
% return values are *not* normalised

% default args
if nargin < 1, obj = target(); end
if nargin < 2, noise = 1; end
if nargin < 3, tile = 3; end
if nargin < 4, sub = 3; end
if nargin < 5, cut1 = 0.5; end
if nargin < 6, cut2 = cut1; end

[h w] = size(obj);

% check pattern big enough to subset for all samples
chk = checkerboard(tile, 1 + ceil(h/(tile * 2)), 1 + ceil(w/(tile * 2))) > 0.5;

% generate phased illumination patterns with OTF (ie, at object)
ff = cell(1, sub * sub);
step = 2*tile/sub;
for x = 1:sub
    xoff = floor((x-1)*step);
    for y = 1:sub
        yoff = floor((y-1)*step);
        ch1 = chk( (1+yoff):(h+yoff), (1+xoff):(w+xoff) );
        ff{1, (x-1) * sub + y} = obj .* otfblur(ch1, cut1);
    end
end

% add constant noise of given intensity and apply second OTF to get image
n = blur(rand(h,w) * noise, 20);
for x = 1:(sub*sub)
    ff{1, x} = otfblur(ff{1, x} + (n/2), cut2);
end

% c0 is uniformly-illuminated version
c0 = otfblur(n + obj, cut2);

% c1 is image reconstructed from frames
c1 = excess(ff);
```

```
function [c0 c1 c2 vv hh] = sincombo ( obj, noise, period, cut1, cut2 )
%SINCOMBO simulates perpendicular excitations with a sin grating
% [C0 C1 C2 VV HH] = SINCOMBO ( OBJ, NOISE, PERIOD, FRAC1, FRAC2 )
%

if nargin < 1, obj = target(); end
if nargin < 2, noise = 1; end
if nargin < 3, period = 6; end
if nargin < 4, cut1 = 0.5; end
if nargin < 5, cut2 = cut1; end


[f1 f2 f3 c0 hh] = sinsim3( obj, noise, period, cut1, cut2);

% transpose for perpendicular version, then transpose back
[f1 f2 f3 x0 vv] = sinsim3( obj.', noise, period, cut1, cut2);
vv = vv.';

c1 = normalise(max(vv, hh));
c2 = normalise(vv + hh);
```

Bibliography

[1] Sebastian Barg, Charlotte S. Olofsson, Jenny Schriever-Abeln, Anna Wendt, Samuel Gebre-Medhin, Erik Renström, and Patrik Rorsman. Delay between fusion pore opening and peptide release from large dense-core vesicles in neuroendocrine cells. 33:287–299, January 2002.

[2] Andrew L. Barlow and Christopher J. Guerin. Quantization of widefield fluorescence images using structured illumination and image analysis software. *Microscopy Research and Technique*, (70):76–84, 2007.

[3] Max Born and Emil Wolf. *Principles of Optics: Electromagnetic theory of propagation, interference and diffraction of light 7th (Expanded) Edition*. Cambridge University Press, 1999.

[4] Ronald N. Bracewell. *The Fourier Transform and its Applications, 3rd Edition*. McGraw-Hill, 2000.

[5] Tobias Breuninger, Klaus Greger, and Ernst H. K. Stelzer. Lateral modulation boosts image quality in single plane illumination fluorescence microscopy. *Optics Letters*, 32(13):1938–1940, July 2007.

[6] S. Chandrasekhar. *Liquid Crystals, 2nd Edition*. Cambridge University Press, 1992.

[7] Peter J. Collings and Michael Hird. *Introduction to Liquid Crystals: Chemistry and Physics*. Taylor & Francis, 1997.

[8] Jose-Angel Conchello and Jeff W Lichtman. Optical sectioning microscopy. *Nature Methods*, 2(12):920–931, 2005.

[9] Forth Dimension Displays. `http://www.forthdd.com/`.

[10] Takashi Fukano and Atsushi Miyawaki. Whole-field fluorescence microscope with digital micromirror device: imaging of biological samples. *Applied Optics*, 42(19):4119–4124, July 2003.

[11] Rafael C. Gonzalez, Richard E. Woods, and Steven L. Eddins. *Digital Image Processing using MATLAB*. Pearson Prentice Hall, 2004.

[12] Joseph W. Goodman. *Introduction to Fourier Optics, 3rd Edition*. Roberts & Company, 2005.

[13] Mats G. L. Gustafsson. Extended resolution fluorescence microscopy. *Current Opinion in Structural Biology*, 9:627–634, 1999.

[14] Mats G. L. Gustafsson. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of Microscopy*, 198(2):82–87, May 2000.

[15] Mats G. L. Gustafsson. Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution. *Proceedings of the National Academy of Sciences*, 102(37):13081–13086, September 2005.

[16] Eugene Hecht. *Optics, 4th Edition*. Pearson Education, 2002.

[17] Rainer Heintzmann. Saturated patterned excitation microscopy with two-dimensional excitation patterns. *Micron*, 34:283–291, 2003.

[18] Rainer Heintzmann and Pier A. Benedetti. High-resolution image reconstruction in fluorescence microscopy with patterned excitation. *Applied Optics*, 45(20):5037–5045, July 2006.

[19] Rainer Heintzmann and Christoph Cremer. Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating. *Proceedings SPIE*, 3568:185–196, 1998.

[20] Rainer Heintzmann, Thomas M. Jovin, and Christoph Cremer. Saturated patterned excitation microscopy—a concept for optical resolution improvement. *Journal of the Optical Society of America A*, 19(8):1599–1609, August 2002.

[21] HoloEye Photonics. `http://www.holoeye.com/`.

[22] Daniel Krueerke. Speed may give ferroelectric LCOS edge in projection race. *Display Devices*, (Fall):29–31, 2005.

[23] Leo G. Krzewina and Myung K. Kim. Single-exposure optical sectioning by color structured illumination microscopy. *Optics Letters*, 31(4):477–479, February 2006.

[24] Jelena Mitić, Tiemo Anhut, Matthias Meier, Mathieu Ducros, Alexander Serov, and Theo Lasser. Optical sectioning in wide-field microscopy obtained by dynamic structured light illumination and detection based on a smart pixel detector array. *Optics Letters*, 28(9):698–700, May 2003.

[25] M. A. A. Neil, R. Juškaitis, and T. Wilson. Method of obtaining optical sectioning by using structured light in a conventional microscope. *Optics Letters*, 22(24):1905–1907, December 1997.

[26] M. A. A. Neil, T. Wilson, and R. Juškaitis. A light efficient optically sectioning microscope. *Journal of Microscopy*, 189(2):114–117, February 1998.

[27] Nikon MicroscopyU. `http://www.microscopyu.com/`.

[28] Costas Pitris and Peter Eracleous. Transillumination spatially modulated illumination microscopy. *Optics Letters*, 30(19):2590–2592, October 2005.

[29] K. Rosenhauer and K.-J. Rosenbruch. The measurent of the optical transfer functions of lenses. *Reports on Progress in Physics*, 30:1–25, 1967.

[30] L. H. Schaefer, D. Schuster, and J. Schaffer. Structured illumination microscopy: artefact analysis and reduction utilizing a parameter optimization approach. *Journal of Microscopy*, 216(2):165–174, November 2004.

[31] Peter T. C. So, Hyuk-Sang Kwon, and Chen Y. Dong. Resolution enhancement in standing-wave total internal reflection microscopy: a point-spread-function engineering approach. *Journal of the Optical Society of America A*, 18(11):2833–2845, November 2001.

[32] Per A. Stokseth. Properties of a defocused optical system. *Journal of the Optical Society of America*, 59(10):1314–1321, October 1969.

[33] Tony Wilson and Colin Sheppard. *Theory and Practice of Scanning Optical Microscopy*. Academic Press, 1984.