

High-density optical data storage

A S van de Nes, J J M Braat and S F Pereira

Optics Research Group

Department of Imaging Science and Technology, Faculty of Applied Sciences, Delft
University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

E-mail: j.j.m.braat@tudelft.nl

Received 3 April 2006

Published 5 July 2006

Online at stacks.iop.org/RoPP/69/2323

Abstract

Since the introduction of optical data storage systems in the 1970s, we have observed a stepwise increase in their storage capacity using the same means for resolution improvement as in classical microscopy and optical lithography, namely, a reduction in the source wavelength and an increase in the numerical aperture of the imaging optics. In this paper we briefly address the historical development of optical data storage and some recent developments towards higher density such as non-linear recording methods and systems with a numerical aperture larger than unity. More specifically, we explore the possibility of storing more information bits per storage location so that optical ‘multiplexing’ becomes feasible. A multiplexing method based on the detection of optical angular momentum of a focused light beam is treated in detail and is illustrated with some examples of preliminary experiments on this subject. Both the existing high-density systems and the proposed new ones require a detailed analysis of the focusing of the scanning spot and the diffraction by the information structure on the disc. We analyse electromagnetic focusing in multilayers and treat the diffraction of light by optical effects using a three-dimensional form of Green’s tensor formalism.

This article was invited by Professor G Leuchs.

Contents

	Page
1. Overview of optical data storage	2325
2. Principles of optical data storage	2327
2.1. Optical transfer	2329
2.2. Modelling of the read-out system	2331
2.3. New high-density systems	2332
3. High-density optical recording	2334
3.1. The illumination	2334
3.2. Interaction with the data	2336
4. Near-field optical recording	2338
5. Multiplexing	2339
5.1. Staircase structures	2340
5.2. Experimental setup	2341
5.3. Angular momentum in the data-structures	2341
5.4. Angular momentum in the light beam	2345
6. Conclusion	2346
Appendix A. Field in the focal region	2348
Appendix A.1. Lens rotation matrix	2349
Appendix A.2. Layer transitions	2349
Appendix A.3. Generalized lens matrix	2352
Appendix A.4. The definition and calculation of Strehl intensity in the high-numerical-aperture case	2353
Appendix B. Solution to the general scattering problem using the Green's tensor formalism	2355
Appendix B.1. Solution of the scattering problem	2355
Appendix B.2. The Green's tensor formalism	2356
Appendix B.3. Stratified media	2358
References	2361

1. Overview of optical data storage

The first thoughts about the storage of data by optical means go back to the 1920s when engineers were looking for new methods to store pictures and sound signals from the newly discovered television medium. Even at that moment in time it was clear that optical methods should be able to realize a much higher storage capacity than the acoustic phonograph. Magnetic storage methods were not yet considered. The oldest patent about optical storage goes back to John Logie Baird, the famous inventor of modern TV [1]. In his UK patent he presented a system with a lamp that focused its radiation on a black-and-white storage medium and the modulated transmitted light was then led to a detector. The storage capacity was enough for 15 min of TV with a bandwidth of typically 20 kHz. Even at that time, such a TV bandwidth yielded unacceptably bad images and his system never made it to the market. Interest in optical data storage was renewed after the success of magnetic storage of TV-images with the professional magnetic tape system of the AMPEX company. Because of its intrinsic complexity and corresponding high price, engineers looked for storage solutions that would give the mass consumer access to data storage in the home. The result was a 'miniaturized' version of the magnetic storage principle, finally resulting in the VHS video recorder standard from the Japanese JVC company. Several other companies explored the possibilities of optical data storage. A first research effort was carried out at Stanford University, sponsored by the 3M-company. In a paper from 1969, Rice *et al* [2] described a mercury-lamp-based optical storage system that used several of today's principles. For instance, storage of the information in a tiny microscopic track was proposed using some modulation method for the electrical signal to be translated into a sequence of optical effects. Due to the thermal light source, neither a good signal-to-noise ratio could be obtained nor an adequate track following achieved. The system was not acceptable for a mass-consumer application. In the beginning of the 1970s, systems were proposed that basically resemble our actual optical data storage systems. The breakthrough was realized thanks to

- the use of a compact low-price helium–neon laser as the light source,
- low-price components while maintaining a high-quality optical system,
- derivation of optical error signals to detect the deviation from optimum read-out conditions (focus, lateral off-set, disc tilt, disc thickness, etc)
- implementation of dynamic control systems for obtaining a correct axial focusing and lateral tracking of the spiral that contains the information and
- development of advanced modulation schemes to optimally adapt the electrical signal to the optical storage medium.

With these advancements, the initial optical players for television signals were launched in the market at the end of the 1970s. Massive market penetration did not happen because the systems were relatively bulky and, not unimportantly, unable to record new information; this was in contrast to the newly introduced magnetic tape systems.

New developments drastically changed the picture. The advent of the compact AlGaAs-semiconductor laser and cheap digital coding and error correction techniques made the new compact disc system feasible. It was devised for audio information of superior quality and immediately became a success after its introduction in 1983, mainly because of the compactness of the disc player and the corresponding standardized discs and its long playing time. Portable versions have followed the initial designs and the material to be stored changed from digital audio to digital in general, enabling its introduction in the personal computer environment. Together with the recording / erasure possibility, the original audio medium became the digital

Table 1. Physical parameters of optical disc systems; storage capacity.

System	Year	λ [μm]	N_O	λ/N_O [μm]	Capacity [GB/layer]	Diameter [cm]	Playing time [minutes]
Video long play	1978	0.633	0.40	1.56	4.5	30	30–60
Laser disc	1983	0.785	0.50	1.57	4.5	30	60
Compact disc	1983	0.785	0.45	1.74	0.65	12	74 (3–4)
DVD	1995	0.650	0.60	1.08	4.7	12	135 (8)
HD-DVD	2006	0.405	0.65	0.62	16	12	135 (12)
Blu Ray	2006	0.405	0.85	0.48	23	12	135 (15)

data storage medium of choice because of its removability that greatly facilitates exchange of information and also because of the extremely low cost per stored bit.

Following the history of optical data storage, one can see that the storage capacity increases with time as presented in table 1. Here, as basic physical parameters of an optical disc system, we have listed the source wavelength λ , the numerical aperture (N_O) of the scanning objective for the various optical disc systems and the corresponding diffraction unit (λ/N_O). The other entries apply to quantities that are of principal importance for the user such as data capacity in gigabyte and playing time and/or data retrieval time (between parentheses), both in minutes. Note that the switch from the relatively bulky helium–neon laser to the tiny semiconductor laser was already made for the analog LaserVision video system. But because of the large-sized discs (30 cm diameter), the benefit to the size of the playing system was negligible. The storage capacity of these first-generation optical discs in bytes can be estimated by taking the bandwidth of the analog system (5 MHz) and multiplying this by 2 (two bits per smallest period of the signal) and by the achieved playing time. The important reduction in size of the system was obtained in the following generations by reducing the disc size to a diameter of 12 cm. A rough estimate of the storage capacity follows the Rayleigh criterion; applied to optical discs, two bits should be spaced apart by a distance equal to $0.61 \lambda/N_O$. With a useful storage area A_S of typically 92.7 cm^2 maximum (with inner and outer recording radii of 23 mm and 59 mm, respectively), the storage capacity C in bytes is given by

$$C = \frac{1}{8} \frac{A_S}{(0.61\lambda/N_O)^2} = 0.34 \frac{A_S N_O^2}{\lambda^2}. \quad (1)$$

Careful inspection of the table shows that the original CD-system approximately adheres to the Rayleigh resolution criterion although the capacity is slightly less than could be expected. The reason is that the system parameters of this first digital optical disc system have been chosen in a rather conservative manner to assure a reliable functioning of the complete system, including mass-produced player and discs. When considering the DVD-system, the Rayleigh criterion prediction for the capacity leads to 2.0 GB, the realized value amounts to 4.7 GB. The reason for this important increase in relative capacity stems from the progress made in digital signal theory and practice (coding, modulation and detection systems, digital hardware) plus acquired experience in mass fabrication and system tolerancing. This same progress in system design and fabrication has also been used in the systems based on the blue GaN laser, operating close to 400 nm wavelength. The same relative increase in density could be maintained for the Blu Ray and HD-DVD systems despite the tighter tolerances on system and components due to the shorter wavelength of the source.

Further developments in optical data storage have been projected on a data storage road map. The road map shows, not surprisingly, that a saturation in system capacity is observed due to the fact that the single-layer capacity is limited by a further decrease in the size of the diffraction unit λ/N_O . A further reduction in wavelength to the UV-region does not seem to be

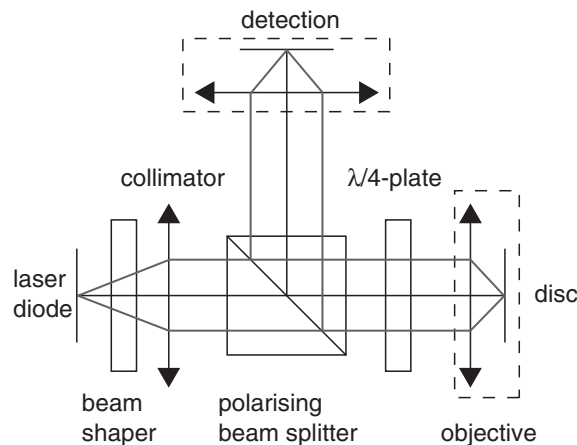


Figure 1. Schematic overview of the optical recording system. The light, emitted by a laser diode, passes a beam shaper, collimator, polarising beam splitter, quarter-wave plate and objective to be focused on the optical disc. The reflected light is collected by the same objective, passes the quarter-wave plate, the polarising beam splitter and the detector lens and is finally imaged on the detector.

realistic at this moment. The only possibility is found in an increase in the numerical aperture beyond unity by using a so-called solid immersion lens (SIL). However, a sensitive system such as this is not yet accepted by the data storage community as it might compromise the easy removability of optical media. Alternative ways to achieve higher capacity rely on the use of nonlinear optical effects at the reading or recording stage or making use of information multiplexing. In this paper, we will concentrate on the latter option where we store more than one bit of information in a single storage location. Apart from increasing the storage capacity, this method also allows a proportional increase in the data retrieval rate. This latter quantity is an important system parameter because the user wants to have a means to quickly retrieve the full data content to another information carrier, be it a magnetic or an optical disc or a high-speed internet link. The retrieval data rate can be increased by spinning the disc at its maximum speed, typically 50 or even 100 Hz rotational frequency. But the retrieval time is increased by the increased number of windings of the information spiral on the disc and this explains the relatively high retrieval time of, e.g. a Blu Ray disc.

The detailed organization of the paper is as follows. In section 2 we briefly review the basic optical principles of an optical disc system. Section 3 is devoted to the modelling of a modern disc player that should be capable to play all existing disc formats listed in table 1. In section 4 we give a description of some new developments to obtain higher capacity optical storage systems, using solid immersion lenses. Finally, in section 5, we treat various multiplexing options and then concentrate on our research to obtain multiplexed information by imparting orbital angular momentum to a focused light beam on reflection at the information track and exploiting the particular far-field diffraction pattern of such a beam. With section 6, the paper terminates with a short summary of the conclusions and a discussion of the future prospects for optical data storage.

2. Principles of optical data storage

A schematic overview of a typical optical recording system is shown in figure 1. The light is emitted by a laser diode, collimated and shaped further to obtain a uniformly distributed light

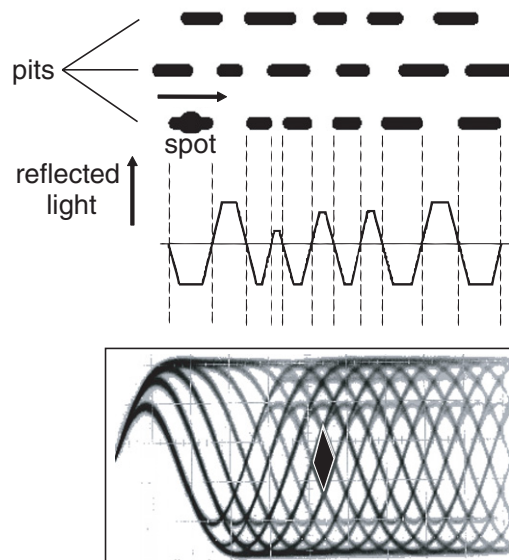


Figure 2. A schematic picture of the data structure on an optical disc. On scanning a track with the focused light spot, a detection signal is found that shows a maximum slope proportional to N_O/λ , the inverse of the lateral size of the spot. The ‘digital eye’, see the drawn lozenge in the figure, becomes visible once a large number of signal sequences is superimposed on, e.g. an oscilloscope that triggers on an arbitrary positive transient in the digital signal; the sequences create a so-called eye pattern. The lateral shift of an individual transient should not exceed half the width of the digital eye to avoid decision errors on the state of the binary signal.

beam with a high quality phase-front. Next, the light is changed to a circular polarization state and focused on the disc. The same objective collects the reflected light which is imaged on the detector.

Nowadays, the signals to be stored on an optical data disc are digital and their quality and robustness at retrieval have to be guaranteed by the optical recording/read-out system. The main factors that determine the quality of a digital signal are the following.

- **Clock regeneration.** The transients in the digital signal create a so-called self-clocking signal that serves as the time base for determining at which moments the binary signal is zero or unity. The minimum distance d in time between two transients determines the run-length of a digital signal. The time increment between two clock ticks gives the minimum length difference between two sequences of zeros (or ones). This increment determines the width of the ‘digital eye’, see figure 2, and should be made as small as possible for obtaining high density.
- **Time jitter.** The recording and read-out mechanism can lead to apparent time shifts of the bits to be detected. Mechanisms that introduce time shifts can be of optical origin such as cross-talk (signal perturbation from adjacent signal tracks) or intersymbol interference (signal perturbation due to the adjacent in-track information bits). The optical recording mechanism, e.g. a local transition from the crystalline to the amorphous phase of the information layer, is based on thermal heating, and thermal diffusion can also be at the origin of apparent shifts of a recorded information bit. If the root mean square value of the time shifts of the transients in the digital signal attains a value of typically 1/6 of the width of the digital eye, errors are likely to occur and the digital signal becomes unreliable.

- Signal-to-noise ratio. While the time jitter gives rise to uncertainty of the bit detection because of a horizontal shift of the transient in the digital eye, a bad signal-to-noise ratio leads to an uncertainty in the detection moment because of the stochastic variations in the signal strength. Especially in multi-level digital signals, the signal-to-noise ratio can be a critical parameter. For two-level signals an overall signal-to-noise ratio of 30 dB is acceptable.

Even if all conditions for correct digital signal reconstruction are satisfied, disc scratches or temporary strong defocus might give rise to full signal loss. For those cases, error correction schemes based on the integrity of larger sequences of bits have been devised. In this way, a relative immunity with respect to in-track perturbations is obtained. However, the solution against track-loss (comparable to the hang-up on old long play audio discs) is much more difficult and requires the use of a larger amount of memory in the optical player.

2.1. Optical transfer

The optical read-out process is most readily described by means of the optical frequency transfer function. This concept is limited to the domain of linear systems and, as such, is not appropriate for the read-out of an optical disc with the single-spot scanning method.

In a seminal paper it was demonstrated by Welford [3] that scanning microscopy is directly comparable to classical microscopy and has to be treated using an adapted version of the general imaging theory for objects that are illuminated with partially coherent light [4]. The theory of the scanning microscope was further developed by Wilson and Sheppard [5,6]. The analogy between the scanning and the classical microscope is easily explained by means of figure 3. The basic resolution in the two systems is provided by the objectives, O_S in the scanning microscope and O_I in the classical microscope. The particular effects of the setting of the condenser aperture C in the classical microscope on the coherence of the object illumination are well described by the van Cittert-Zernike theorem [4]. The same coherence effects can be obtained in the scanning microscope by adapting the size of the aperture in front of the detector D . In both cases, it is the sine of the opening angle of the condenser or the detector that determines the 'partial' coherence, denoted by the factor σ . An opening angle approaching zero leads to fully coherent illumination or detection ($\sigma = 0$). If the sine of the opening angle of the condenser or detector is much larger than the objective aperture, the illumination can be said to be fully incoherent ($\sigma \rightarrow \infty$). In the case of an optical disc system used in reflection, a natural choice is an equal aperture for objective (N_O) and detector (N_D) which leads to a partially coherent setting that is close but not equal to the fully incoherent case. The resulting extreme cases in spatial frequency transfer from the scanned object to the detector are illustrated in figure 4. The practical solution adopted in optical disc systems $N_D/N_O = 1$ asks for a treatment that is linear neither in optical amplitude (coherent case, $\sigma = 0$) nor in optical intensity (fully incoherent case, $\sigma \rightarrow \infty$). The subtle differences between the various stages of 'coherence' in the read-out of an optical disc is best illustrated by taking a one-dimensional binary object (step function) that produces an abrupt change in reflectivity from zero to unity. The intensity response $S_D(x, y)$ of the detector if the scanning spot is exactly centred on the transition ($x = 0$) is given by

$$S_D(0, y) = \begin{cases} 1/4 & (\sigma = 0) \\ 1/3 & (\sigma = 1) \\ 1/2 & (\sigma \rightarrow \infty) \end{cases} . \quad (2)$$

The upper and the lower results can be easily derived by summing respectively the amplitude or the intensity of the scanning spot that is 50% transmitted or reflected by the step function.

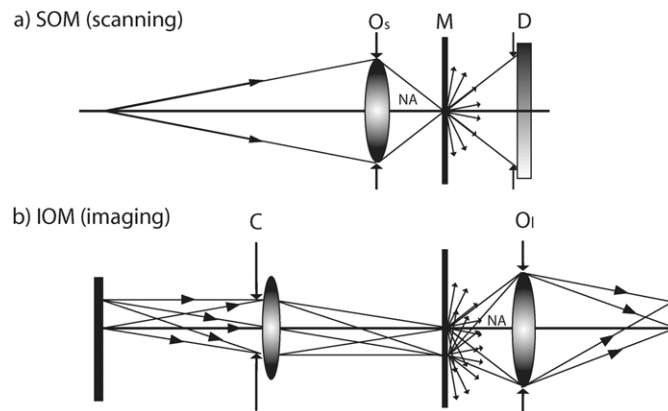


Figure 3. (a) The scanning optical microscope SOM and (b) the ‘classical’ imaging optical microscope IOM with simultaneous illumination of the full field (as an example, two source points out of the extended incoherent source are shown in the drawing). The scanning microscope is optically characterized by the numerical aperture of its objective and the collecting angle (aperture) of the detector (dashed cone). In the figure, the aperture N_D of the detector has been taken equal to the numerical aperture N_O of the objective O_S , a situation that corresponds to standard reflective read-out of an optical disc.

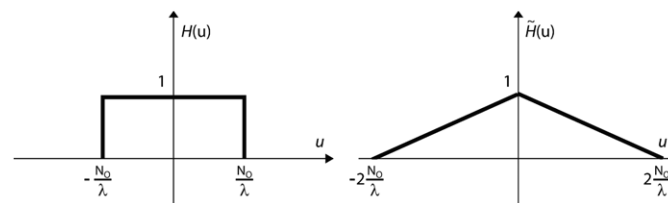


Figure 4. Schematic drawing of the modulation transfer function as a function of the spatial frequency in the scanned object. The function $|H(u)|$ corresponds to the fully coherent case (left-hand graph) and has a high-level uniform frequency transfer up to a limiting frequency equal to N_O/λ with a the half diameter of the objective pupil and f_1 the focal distance. In the fully incoherent case ($\sigma = N_D/N_O \gg 1$) the frequency transfer ($|\tilde{H}(u)|$) is nonzero up to twice as large value but the modulation of the detected signal is continuously decreasing for larger spatial frequency u .

The result for $S_D(0, y)$ in the partially coherent case with $\sigma = 1$ is more intricate; an analytical derivation of the result $1/3$ was given long ago in [7].

The conclusion from the diffraction calculations above is that fully incoherent detection ($\sigma = \infty$) produces highest liability in detecting transients of the information track. Because of the impossibility to realize this value in practice, optical data detection has to take into account the effects of finite ‘coherence’ and accommodate lateral shifts of transients of the order of $0.1 \lambda/N_O$. The shifts are partly predictable by prior knowledge about the bit sequences to be detected using modulation schemes but, for a substantial part, these shifts are of a more random nature once the influence of neighbouring tracks is also incorporated (cross-talk). For this reason, a basic uncertainty in transient position of typically $0.03 \lambda/N_O$ is unavoidable in optical data storage; for the Blu Ray system, this value amounts to approximately 10–15 nm, to be compared with a basic feature length of 150 nm. Higher density on optical discs is hampered by this basic optical uncertainty, together with the influence of noise sources such as electrical noise, laser noise, photon shot noise and, last but not least, medium noise.

2.2. Modelling of the read-out system

The rather superficial analysis of the preceding paragraph gave a global idea of the limits that are encountered in achieving higher density. Note that in this paper we do not focus on the recording process (optical disc mastering, phase-change recording or magneto-optic recording); we suppose that means are available for producing optical effects with a certain density. Our emphasis here is on a more detailed modelling of the read-out process that goes beyond the standard approach described in the earlier literature. The first publications on optical disc systems used a relatively simple optical scattering model. An early experimental system was based on the use of very high resolution photographic material [2]. Instead of this, because of ease of replication, relief structures were proposed and the design of the optical ‘pits’ was based on achieving a phase difference of π between the light reflected from a depression (or ‘bump’) and the light reflected from the environment (‘land’ region) [8–11]. Later publications used a more elaborate model, analogous to the scanning microscope model; the disc structure is still represented by an infinitely thin reflection screen with a spatially varying complex reflection coefficient [12–16]. A full account of the scanning microscope approach to optical disc read-out is found in [17] and this paper also gives a detailed description of the numerical implementation of the corresponding scalar diffraction theory. A system description using this same scalar diffraction picture of optical disc read-out that turned out to be sufficiently accurate for the lower density optical discs systems such as laser disc, CD and DVD can be found in [18, 19]. General descriptions of optical disc systems are also found in [20–28].

2.2.1. Vector focusing. At relatively low numerical aperture, the calculation of the field in the focal region is carried out using scalar diffraction theory. At higher numerical aperture one has to include the state of polarization of the light as was done in early references on this subject [29–32]. Further refinements include the presence of aberrations, either of circularly symmetric nature [33, 34] or of a general nature [35–38], and the incorporation of several layer transitions in [39]. As the geometrical numerical aperture in the information layer of an optical disc has been reduced by a factor of $\text{Re}(n_{\text{med}})$ (the real part of the refractive index of the layer), the vector effects in focus remain relatively small, even for the high numerical aperture Blu Ray system. However, as soon as the solid immersion lens (SIL) method of focusing is introduced [40–42], the vector effects become very pronounced and have to be included in order to obtain a good prediction of the detector signal retrieved from the high-density disc.

2.2.2. Vector diffraction. At an early stage it was recognized that the scalar diffraction theory was basically insufficient to accurately account for all observed effects in optical disc read-out. From the very beginning, the lateral dimensions of an optical effect were of the order of the wavelength of the light and this means that the electromagnetic boundary conditions used in, for example, Kirchhoff’s scalar diffraction approach are inadequate. The influence of the state of polarization on the diffracted far-field and the apparent depth of optical pits as a function of incident field polarization was described in [43] using the numerical methods for electromagnetic diffraction that were available at that moment [44–46]. Later research has aimed at an improvement in convergence and speed of these three-dimensional diffraction calculation methods based on plane wave incidence. A more recent treatment using a rigorous two-dimensional diffraction theory is given in [47, 48]. High-numerical-aperture read-out of an optical disc pattern needs information about the scattered field for a large number of angles of incidence of the elementary plane wave and this leads to a high computational load.

2.3. New high-density systems

In this subsection we briefly describe potential high-density systems in optical data storage that could find a position in the mass-consumer market in the coming decade. We start by excluding a certain type of system, namely, the holographic page-oriented systems. These systems are based on classical holography [49, 50] and nowadays use angular multiplexing to store images or bit patterns and to retrieve them from a CCD-type array detector. More recent systems use specific properties of volatile recording materials such as bacteriorhodopsin and rely on polarization holography [51, 52]. In all cases, the holographic systems lack the low-price capability that is needed for a penetration of the mass-consumer market. The holographic systems also lack an easy replication option which makes them unsuited for distribution of huge amounts of data. For this reason we will not consider volumetric recording systems in this paper and concentrate on systems that achieve the highest possible capacity per individual information layer. In general, a single layer disc can be extended to a multi-layer system as has been done in the DVD-standard. The high-density systems based on nonlinear effects are only briefly considered; they exploit temperature-induced nonlinear refractive index effects [53] and/or super-resolution effects in, e.g. crystalline/amorphous thin layers based on antimony or silveroxide (super-RENS [54, 55]). Other systems use magneto-optic thin layer stacks (MAMMOS [56]). The complexity or limited durability of these nonlinear information stacks seems to prohibit them from becoming a widely accepted standard and from penetrating the market.

To illustrate the effect of typical nonlinear read-out behaviour we suppose the following dependence of the refractive index on the focal spot intensity $I(x, y)$

$$n(x, y) = n_0 + \Delta n I(x, y) = n_0 + \Delta n |a(x, y)|^2, \quad (3)$$

where $a(x, y)$ represents the complex amplitude of the electric field in the focal region.

In the simplified scalar diffraction approximation, the reflected amplitude $a'(x, y)$ in the immediate 'near-field' is given by the product of the reflection function, approximated by that of a thin phase object, and the incident wave function leading to

$$a'(x, y) = \exp [i2k_0 (n_0 + \Delta n |a(x, y)|^2) h(x, y)] a(x, y), \quad (4)$$

where k_0 equals the wave number in vacuum and $h(x, y)$ describes the height profile of an information pit on the optical disc.

A height profile with the property $h(x, y) \ll \lambda$ leads to the approximate expression

$$a'(x, y) \approx [1 + i2k_0 (n_0 + \Delta n |a(x, y)|^2) h(x, y)] a(x, y). \quad (5)$$

The complex amplitude $B(X, Y)$ reflected back into the exit pupil of the scanning objective is given by the Fourier transform of $a'(x, y)$ and we obtain

$$B(X, Y) \approx A(X, Y) + 2ik_0 n_0 H(X, Y) * A(X, Y) + 2ik_0 \Delta n H(X, Y) * FT [|a(x, y)|^2 a(x, y)], \quad (6)$$

where the functions $A(X, Y)$ and $H(X, Y)$ are the Fourier transforms (FT) of $a(x, y)$ and $h(x, y)$, respectively, and (X, Y) are the normalized Cartesian co-ordinates of a general point on the pupil sphere.

Using the standard detection method employed in optical disc players, one collects all the light captured within the exit pupil of the objective on the way back from the disc. The detector current is then proportional to

$$i_D \propto \int \int_{S_{\text{pupil}}} |B(X, Y)|^2 dXdY. \quad (7)$$

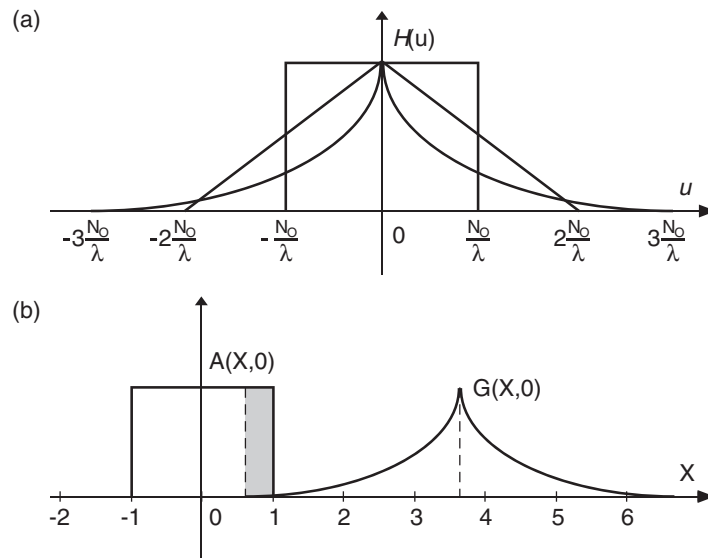


Figure 5. (a) A schematic drawing of the Fourier transform $G(X, Y)$ of the function $g(x, y) = |a(x, y)|^2 a(x, y)$ as compared with the FTs of $a(x, y)$ and $|a(x, y)|^2$, plotted in terms of normalized transfer functions $H(u)$ (upper figure). (b) A cross-section in the plane $Y = 0$ of the function $G(X, Y)$, off-set by a x -frequency component close to $4N_O/\lambda$ and its contribution within the detection pupil (shaded region within the detection pupil $A(X, 0)$ with normalized extent ± 1).

In the case of a perfect diffraction-limited scanning spot $a(x, y)$, its Fourier transform $A(X, Y)$ simply reduces to the so-called top-hat function, equal to unity within the pupil of the objective and zero outside. In the absence of any information on the disc, the function $H(X, Y)$ is a δ -function and the contribution from the second term in equation (6) reduces to $2ik_0 n_0 A(X, Y)$. The presence of the optical information pits leads to a spatial modulation of the function $h(x, y)$ and the Fourier transformed function $H(X, Y)$ acquires a certain extent. Part of that spatial extent might not be contained within the detection solid angle determined by the exit pupil size and may be lost for detection purposes. But the convolution product $H(X, Y) * A(X, Y)$ leads to a broadening with respect to the original function $H(X, Y)$ so that the information stored in the tracks on the disc can still be captured by the detector. A careful analysis shows that spatial frequencies corresponding to a sine of the diffraction angle that is twice the value of the numerical aperture of the objective are still present in the detector signal, albeit with a strength that approaches zero, see figure 4. The influence of the nonlinear effect, present in the third term of equation (6), is a further broadening of the frequency transfer due to the convolution with the Fourier transform of the function $g(x, y) = |a(x, y)|^2 a(x, y)$. In figure 5 we have plotted the basic shape of the function $G(X, Y) = FT\{g(x, y)\}$. It is seen that the total support of this function is $\pm 3N_O/\lambda$. This means that frequency components in $H(X, Y)$ close to $4N_O/\lambda$ can be folded back and still be transmitted through the outer part of the exit pupil of the objective. However, in practice, the nonlinear effect due to, e.g. thermal heating in phase-change recording layers is small and the resulting signals are weak. A negative side-effect is that the repeated heating of a layer tends to be destructive for the stored information.

An improved version of the nonlinear read-out described above is obtained by the so-called RENS or super-RENS principle [54, 55]. With an extra absorbing layer showing a pronounced nonlinear behaviour, the read-out spot is reduced in size by means of the local thermal bleaching of the nonlinear auxiliary layer. The corresponding sharpening of the spot

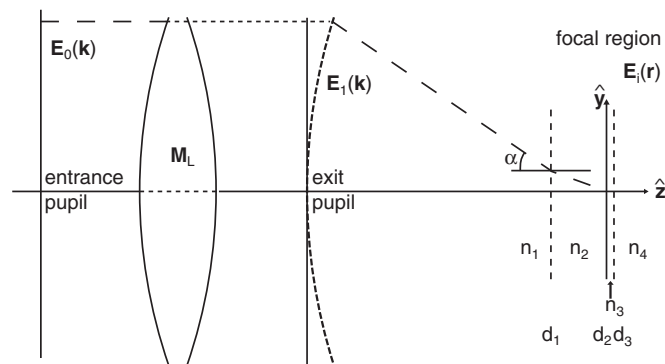


Figure 6. A circularly polarized light beam with a uniform amplitude and phase distribution is imaged on the optical disc in the focal region. The operator M_L represents the objective with $N_O = \sin \alpha$, rotating the electric field vector in the entrance pupil to be transversal to the spherical cap with as its origin the focal point. The material transitions in the focal region take into account the cover layer (2), the data containing layer (3) and the substrate (4).

leads to a broader band of transferred frequencies, very much like the nonlinear mechanism described above. Fine-tuning of the read-out power is needed to achieve the optimum bleaching window for the reduction of the scanning spot size. Practical problems (stability of the RENS-layer, laser power control, photon efficiency) have so far prohibited the application of this nonlinear recording and read-out principle in optical data storage.

We conclude this section by introducing the potential high-density systems that we will treat in more detail in the rest of this paper:

- systems with a numerical aperture larger than unity (solid immersion lens or ‘SIL’-option)
- multiplexed optical data systems.

Regarding the systems with a numerical aperture exceeding unity and the multiplexed optical systems, we will show how the basic design of the read-out beam and the information structures has to be derived from rigorous calculations that take the vector character of the light into account.

3. High-density optical recording

Since the introduction of the optical disc system, the commercially available optical recording products have advanced to the stage of the Blu Ray disc system. This system approaches the limits of the validity of the scalar theory used to describe the illumination. For an accurate treatment of such a high numerical aperture system, it is necessary to use a theoretical description based on the fully vectorial Maxwell equations, not only for the interaction with the data but also for the modelling of the illuminating scanning spot. We separate the optical recording setup in three different regimes, the illumination of the disc, the interaction with the data and the detection. In this section, we compare a scalar and a vectorial treatment of the current high numerical aperture system, concentrating on the illumination and the interaction with the data since the detection part is still accurately described by a scalar treatment.

3.1. The illumination

The imaging system used for optical recording is effectively similar to a scanning microscope; a schematic overview is given in figure 6. The circularly polarized light beam is focused by the

objective onto the optical disc. The objective has a numerical aperture $N_O = 0.85$, and although the geometrical aperture inside the storage medium has a value of $N_O/n_{\text{med}} = 0.85/1.6 = 0.531$, the light distribution in the focal region shows small but significant differences when calculated using the scalar equations as compared with a fully vectorial description. The Ignatowsky [29] or Richards and Wolf integral [30], derived from the Debye integral [57], is given by

$$\mathbf{E}(\mathbf{r}) = -\frac{i}{2\pi} \int \int_{\Omega} \frac{a(k_x, k_y)}{k_z} e^{i\mathbf{k}\cdot\mathbf{r}} dk_x dk_y, \quad (8)$$

where the integration takes place over the exit pupil of the imaging system, in terms of the spatial-frequency co-ordinates k_x and k_y . The field in the exit pupil is effectively decomposed in plane waves with frequencies k_x , k_y and $k_{z1} = (k_1^2 - k_x^2 - k_y^2)^{1/2}$. The relation of the exit pupil distribution to the entrance pupil distribution is given by $\mathbf{a}(k_r, k_\phi) = R_f(k_{z1}/k)^{1/2} \mathbf{M}_L \cdot \mathbf{E}_0(k_r, k_\phi)$ with R_f the focal length of the lens system and the general lens matrix incorporating possible layer transitions, as derived in [39] and summarized in appendix A,

$$\mathbf{M}_{L,i}^{\pm}(k_r, k_\phi) = \begin{pmatrix} g_i^{0\pm} - g_i^{2\pm} \cos 2k_\phi & -g_i^{2\pm} \sin 2k_\phi & 0 \\ -g_i^{2\pm} \sin 2k_\phi & g_i^{0\pm} + g_i^{2\pm} \cos 2k_\phi & 0 \\ -g_i^{1\pm} \cos k_\phi & -g_i^{1\pm} \sin k_\phi & 0 \end{pmatrix}, \quad (9)$$

where $g_i^{j\pm}$ have been introduced as generalized reflection and transmission coefficients in equation (23). As long as the entrance pupil distribution satisfies the scalar wave equation, the vectorial electric field distribution obtained with this formalism is a solution of Maxwell's equations, as demonstrated in [58]. In the case of a completely scalar situation, i.e. when the numerical aperture is small enough, the integral in a homogeneous medium reduces to

$$U(r, \phi, 0) = -ik_1 R_f N_O^2 E_0 \frac{J_1(N_O k_1 r)}{N_O k_1 r}. \quad (10)$$

Now, we compare the field in the focal plane, calculated with the scalar and the vectorial theory. In figure 7(a) the field in the focal plane is shown for the modulus of the total electric field, $|\mathbf{E}|$, for the transversal field component $|E_t|$ and for the longitudinal field component $|E_z|$. The field as calculated with the scalar theory is, for the relatively small geometrical aperture, almost equal to the transversal component. However, a considerable amount of the total intensity is available in the z component of the electromagnetic field. The difference in the scalar and vectorially calculated intensity profiles in the focal plane are shown in figure 7(b). The total field distribution in the focal region is shown in figure 7(c).

Note that the imaging system has to be corrected for the spherical aberration caused by the cover layer of the system. Due to the layer transition $d_1 = 100 \mu\text{m}$ in front of the focal plane, the field at the focal plane is severely distorted. In a practical situation, the objective corrects for these aberrations, which can be accurately described by incorporating the inverse effect in the lens operator \mathbf{M}_L . The explicit transmission coefficients for the layer transition, see equation (A.21), are given by

$$f_2^{s+}(k_x, k_y) = \frac{2k_{z,1}}{k_{z,1} + k_{z,2}} \exp[i(k_{z,1} - k_{z,2})d_1], \quad (11a)$$

$$f_2^{p+}(k_x, k_y) = \frac{k_2}{k_1} \frac{2\epsilon_1 k_{z,1}}{\epsilon_2 k_{z,1} + \epsilon_1 k_{z,2}} \exp[i(k_{z,1} - k_{z,2})d_1]. \quad (11b)$$

For the design of the objective, only the exponential function present in these coefficients is taken into account. The effect of the reduction in amplitude of the plane waves contributing to the field in the focal region for the higher frequencies is not corrected for and added with

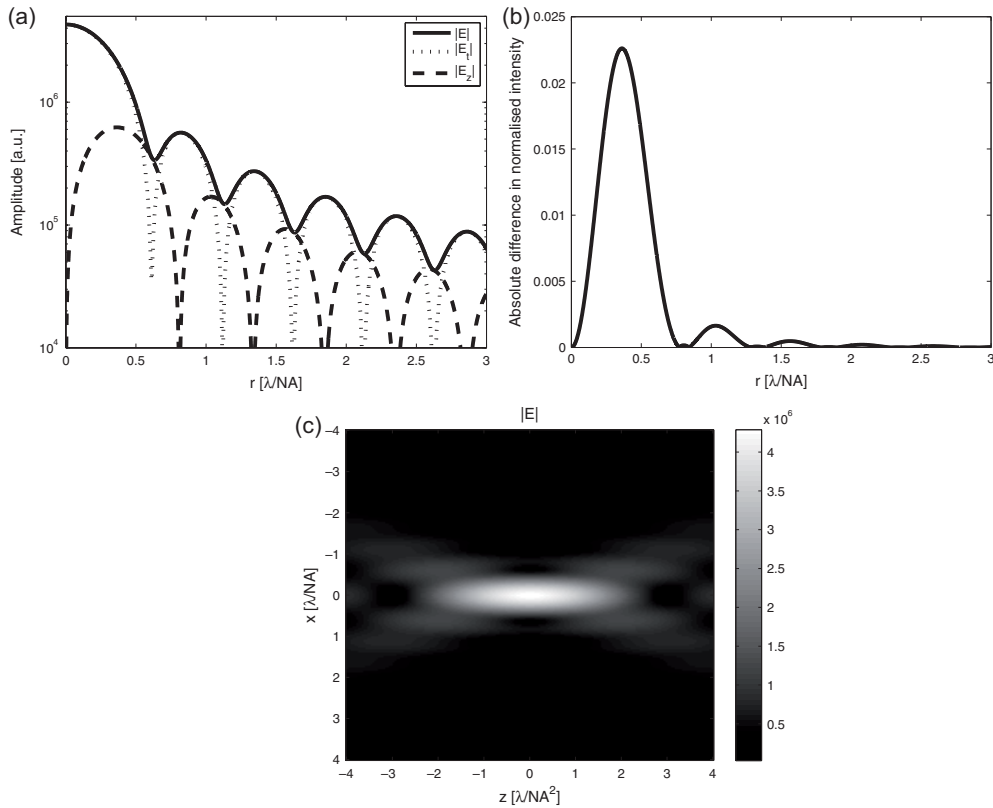


Figure 7. (a) Modulus of the total electric field $|E|$ (—) of the transversal field component $|E_t| = (|E_x|^2 + |E_y|^2)^{1/2}$ (⋯⋯⋯) and the longitudinal field component $|E_z|$ (- - -) in focus for an imaging system with $N_O = 0.85$ and initially a circular polarization state. (b) Difference between the vectorially and scalarly calculated intensities. (c) Modulus of the of the total electric field distribution in the focal region at the (x,z) -plane.

the radiometric effect yields a larger illumination spot. In the low-numerical-aperture case, the optimization of the illumination spot is generally done by maximizing the so-called Strehl intensity using the scalar diffraction model. In [appendix A.4](#), we briefly describe the modified definition of Strehl intensity in the case of a high numerical aperture illumination spot.

The imaging system used for current optical recording effectively consists of four layers. First the air medium in which the objective is located, next the cover layer, the data containing layer and the substrate, with refractive indices n_1 , n_2 , n_3 and n_4 , respectively. To incorporate the effect of these layer transitions, we have used the method as described in [appendix A](#) for calculating the field in the focal region.

3.2. Interaction with the data

Once the field in the focal region is correctly described, the interaction with the scattering structures can be calculated. As stated before, the structures on the disc are of the order of the wavelength. Therefore, for an accurate description, the field reflected by the disc should be calculated with rigorous simulation tools. However, since the emersion of optical recording this was not feasible due to the lack of computer power, so the previously mentioned Kirchhoff

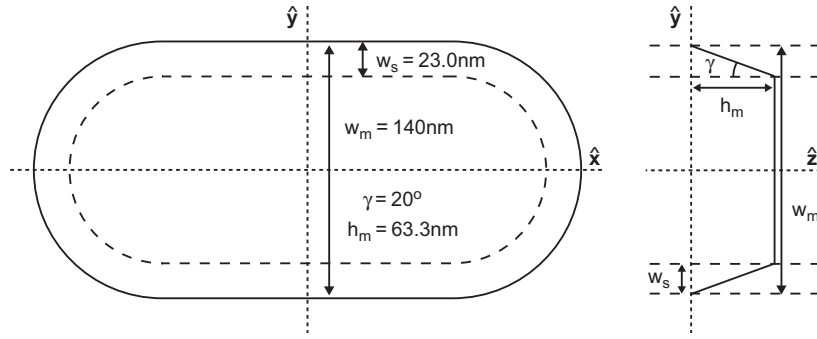


Figure 8. Schematic overview of the pit structure in the metallic layer. The track direction is along the x -axis. The width of the pit, perpendicular to the track direction, is $w_d = 140$ nm, the depth is $h_m = 63.3$ nm, and the slope of the pit is $\gamma = 20^\circ$ which yields a width of the sloping wall of $w_s = 23.0$ nm.

approximation was used, instead where the field in the focal plane is multiplied with a spatially dependent reflection function describing the effect of the data containing structures. Since then these structures have been reduced even further in dimensions, but the computer power has significantly increased; nowadays it is possible to calculate the reflection from the scattering structures on the disc rigorously.

The calculated structures have been schematically depicted in figure 8, where the depth of the structures in the focal plane is described by $h(x, y)$. For a representation of the pit structure, a simplified model has been used following the Blu Ray disc system. At the top of a pit, the width is equal to its design width $w_m = 140$ nm, but at the bottom of a pit, at a depth $h_m = 63.3$ nm corresponding to a quarter-wavelength in the medium n_2 , the width is reduced by twice the factor $w_s = 23.0$ nm corresponding to the width of the sloping wall of the pit with slope angle $\gamma = 20^\circ$. The quarter-wavelength depth of the pit applies only to the ideal value for the scalar situation; in practice, the depth is optimized for obtaining the best contrast. The spatially varying complex reflection function is given by

$$r(x, y) = f(x, y) \exp [ih(x, y)], \quad (12)$$

where the function $f(x, y)$ describes the strength of the reflection in the focal region. Now, the reflected near-field is obtained by multiplication of the function $r(x, y)$ with the field in the focal region. Consecutively, the reflected field at the exit pupil can be obtained by performing a Fourier transform of the reflected near-field. In our simulation we use a homogeneous strength of the reflection function $f(x, y) = 1$.

For a rigorous treatment of the scattering problem, more advanced calculation techniques have to be used, such as finite difference time domain calculations, Fourier modal methods or volume integral methods. We will concentrate on the latter, using a Green's tensor technique to obtain the reflected field in the focal region, as discussed in more detail in [appendix B](#). This method discretizes the geometry of the scatterer and depends on the analytically known response multiplied by the difference of a spatially varying refractive index function with respect to a refractive index background. The total electric field $\mathbf{E}_t(\mathbf{r})$ is given by the sum of the initial electric field $\mathbf{E}_u(\mathbf{r})$ and an electric Green's tensor response \mathbf{G}_e multiplied with the total electric field and the refractive index difference function,

$$\mathbf{E}_t(\mathbf{r}) = \mathbf{E}_u(\mathbf{r}) + \int [\epsilon_t(\mathbf{r}') - \epsilon_u(\mathbf{r}')] \mathbf{G}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{E}_t(\mathbf{r}') d\mathbf{r}', \quad (13a)$$

$$\mathbf{H}_t(\mathbf{r}) = \mathbf{H}_u(\mathbf{r}) + \int [\epsilon_t(\mathbf{r}') - \epsilon_u(\mathbf{r}')] \mathbf{G}_h(\mathbf{r}, \mathbf{r}') \cdot \mathbf{E}_t(\mathbf{r}') d\mathbf{r}', \quad (13b)$$

where the Green's tensors are given by

$$\mathbf{G}_e(\mathbf{r}, \mathbf{r}') = \frac{\kappa^2}{\epsilon} \left[\frac{i\kappa R - 1}{\kappa^2 R^2} (\mathbf{I} - 3\hat{\mathbf{R}}\hat{\mathbf{R}}) + (\mathbf{I} - \hat{\mathbf{R}}\hat{\mathbf{R}}) \right] g(\mathbf{r}, \mathbf{r}'), \quad (14a)$$

$$\mathbf{G}_h(\mathbf{r}, \mathbf{r}') = -i\omega \frac{i\kappa R - 1}{R} (\hat{\mathbf{R}} \times \mathbf{I}) g(\mathbf{r}, \mathbf{r}') \quad (14b)$$

and the Green's function by

$$g(\mathbf{r}, \mathbf{r}') = \frac{e^{i\kappa|\mathbf{R}|}}{4\pi|\mathbf{R}|}. \quad (15)$$

The relative distance between observation and source point is given by the vector $\mathbf{R} = \mathbf{r} - \mathbf{r}'$. As usual, a hat denotes normalization and $|\mathbf{R}|$ denotes the length of the vector. The vectorial product of $\hat{\mathbf{R}}\hat{\mathbf{R}}$ yields a tensor of rank 2 with elements $R_i R_j / |\mathbf{R}|^2$. For our simulation, we use a hypothetical refractive index of the metal layer containing the pits of $n = 5 + 5i$.

In figure 9, we show the difference in the calculated reflected light, where the illumination spot is scanned over the disc, for the Kirchhoff approximated result and the rigorous simulated results (Green's tensor). Note the small modulation depth of the signal, which is still sufficient owing to the high quality electronics post-processing the data. The reduction in modulation depth is partly due to the high numerical aperture read-out; the reflected light that can be scattered outside the detection pupil is a relatively small fraction of the total light power.

4. Near-field optical recording

A logical step to increase the data capacity for the optical recording system is to use an objective lens with an even higher numerical aperture, i.e. beyond the maximum of unity in air. However, this maximum is determined by the refractive index of the material in which the light propagates. In this section, we discuss an example of a solid immersion system which has a $N_O = 1.5$, where the medium of the objective has a refractive index of $n_{\text{sil}} = 2.086$. For a successful commercial product, it is very important to keep the recording medium exchangeable in the system, which means that the objective and the disc should not be permanently connected (removability). To ensure this, the light from the solid immersion lens will have to travel through a sub-wavelength air-gap in order to 'connect' the medium of the objective lens to that of the optical disc. Since the numerical aperture cannot be higher than unity inside the air-gap, higher frequencies are basically not supported and, as a consequence, the light at these frequencies will be reflected. However, the light at these high frequencies can be transmitted as evanescent waves. In [39], we have shown that even for a relatively large air-gap of $\lambda/2$ the lateral size of the spot illuminating the data-layer is still reduced in width. In practice, it is possible to keep the air-gap much smaller at approximately 30 nm, while the focus actuator is still able to obtain a good focus error-signal [42, 59]. The geometrical aperture in the recording medium (polycarbonate) is $N_O/n_{\text{med}} = 1.5/1.6 = 0.94$ and clearly requires a vectorial treatment.

For the solid immersion setup, a relatively thin cover layer is used with still enough thickness to protect the disc from scratches. In our example, we have chosen a thickness of $3 \mu\text{m}$. Again, the lens is expected to correct the phase front for both the cover layer and the air-gap ($\exp[i(k_{z,1} - k_{z,2})d_1 + i(k_{z,2} - k_{z,3})d_2]$). In figure 10, we show the field in the focal region for the solid immersion setup for the total electric field, the transversal component and the longitudinal component. The maximum contribution of the z component to the total electric field is 31%, causing a broadening of the spot in the focal plane. In [42], it is shown that even a four-layer recording system is feasible for near-field recording. Data storage systems based on

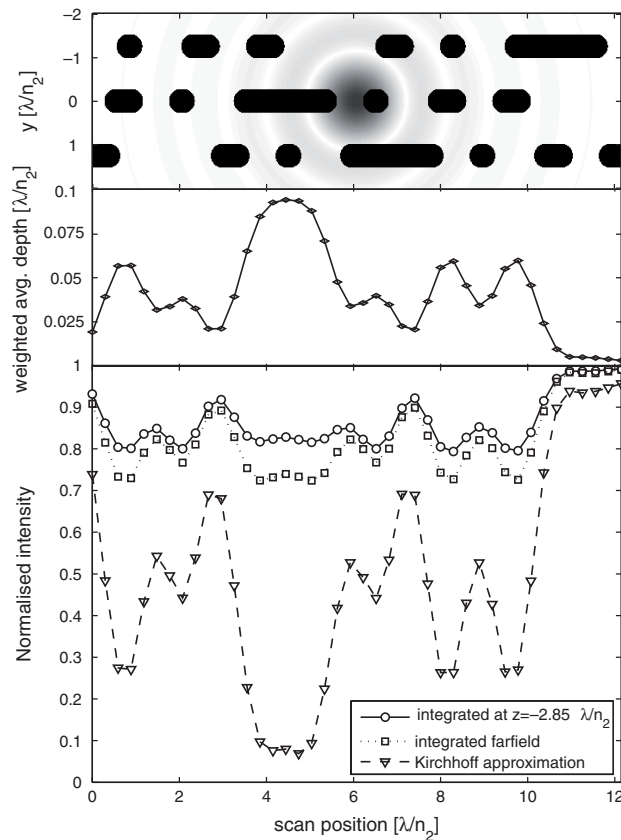


Figure 9. The integrated reflected intensity as a function of a scan along the track (bottom graph) for the scalar results (∇) and the vectorially calculated results $2.85\lambda/n_2$ above the structure (\circ) and in the far-field (\square). The data structures are shown in the top graph, with an example position of the scanning spot in the centre. The middle graph shows the average depth of the illuminated area, weighted for the intensity distribution of the spot.

this principle can be expected to achieve a single-layer capacity of 125 GB and a single-sided capacity of 500 GB.

An alternative for obtaining a circularly symmetric spot is to use radial polarization [39,61], which yields for high numerical aperture systems a smaller transversal width of the illumination spot. The radial and azimuthal polarization states can be obtained by the insertion of a specially designed liquid crystal cell in the beam [62].

5. Multiplexing

In optical recording, the reflected total electric intensity is measured, while scanning the focal spot along the track. Consequently, the shape of the pit in the track direction is translated in terms of a set of frequencies. However, the spatially varying electric field distribution at the detector plane has more information available about the illuminated structure, which is stored in the spatially resolved amplitude, phase and polarization values. Of course, it should be possible to store more information per storage unit if one would detect these variables. This is denoted as information multiplexing. So far multiplexing for optical recording has only been attempted in grey-level recording [60] (multiplexing in amplitude). However, other systems

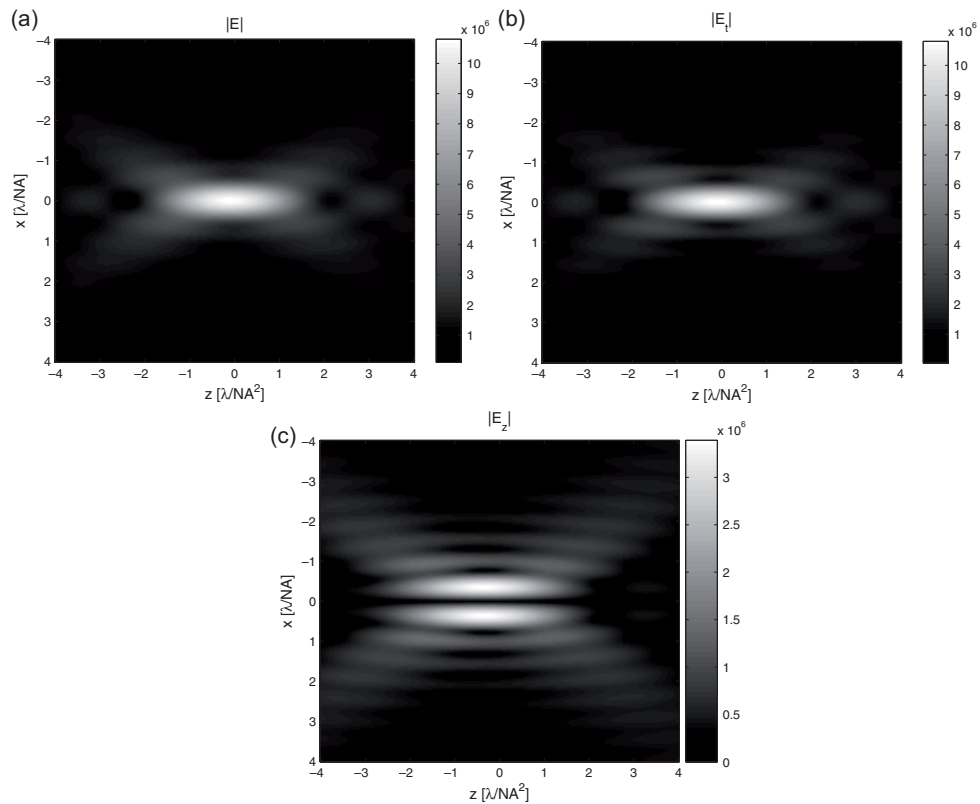


Figure 10. Modulus of the electric field in the focal region for (a) the total field, (b) the transversal component and (c) the longitudinal component for an imaging system with $N_O = 1.5$ and an initially circular polarization state.

have been proposed based on multiplexing in wavelength and frequencies in the direction orthogonal to the scan direction on the disc [63]. These systems have not yet resulted in a substantial gain of the storage capacity. However, multiplexing is only going to be possible when the detected quantities are more or less independent. Here we present a candidate for multiplexing of such an independent variable, angular momentum, available as the phase and polarization information in the detector plane.

To determine whether it will be possible to use angular momentum as a data-channel for optical recording, we performed an experiment using the ‘staircase’ structures discussed in the next section. In the experiment the structures are illuminated by a diffraction-limited focused beam with its centre aligned with the centre of the staircase structures. The structures are imaged on the camera, and each individual image of the obtained series has a different amount of defocus on the camera. The structures transfer orbital angular momentum to the light beam, which is made visible by defocusing the image. For a better understanding of the obtained measured data, we compare the data with numerical calculations based on a decomposition in Gauss–Laguerre modes.

5.1. Staircase structures

To obtain a sense of rotation in the light beam, a set of staircase structures has been fabricated, see figure 11. These structures yield per step a reflected field which is delayed in time with

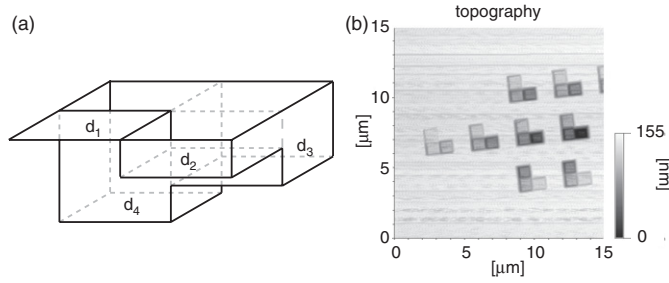


Figure 11. Quadrant staircase structures of dimensions $2\ \mu\text{m} \times 2\ \mu\text{m}$. (a) Schematic overview and identification of the step-size. (b) Topography of the structures obtained with an atomic force microscope.

respect to the previous step. In the limit of infinitely small steps, the rotational sense becomes continuous and approaches a spiral phase-plate [64]. In the case of a finite amount of discrete steps a modal decomposition over the entire set of helical functions is performed. For our experiment, staircase structures with 4 steps were fabricated in a silicon wafer with a focused ion beam (FIB), each structure having a total dimension of $2\ \mu\text{m} \times 2\ \mu\text{m}$. A single step height d_2 is chosen such that $d_1 = 0$, $d_3 = 2d_2$ and $d_4 = 3d_2$. After reflection the various step sizes d_2 are equivalent to phase differences of $\pi/2$, $3\pi/8$, $\pi/4$ and $\pi/8$ for a wavelength of 650 nm.

5.2. Experimental setup

For an experimental verification of the presence of angular momentum, a simple optical interferometer has been used, as schematically shown in figure 12. For illumination of the sample we use a stabilized helium–neon gas laser with $\lambda = 633\ \text{nm}$ and an output power of 1 mW. The emitted light propagates through a beam expander, followed by an optical isolator. Next, the beam is focused by a low numerical aperture objective at 18 cm in front of a microscope objective ($N_O = 0.5$) used for imaging of the sample. Since the structures are slightly larger compared with the helium–neon laser wavelength, the staircase structures are illuminated by a diffraction-limited spot, where the illumination did not completely fill the microscope objective, to effectively yield a numerical aperture of $N_O = 0.4$. This light path forms one branch of the interferometer, the other branch provides a reference beam for alignment purposes. An image of the sample is formed with a magnification of $M \approx 29$, and another microscope objective with a numerical aperture of $N_O = 0.45$ collimates the light. Finally, a low numerical aperture lens re-images the sample on the CCD-camera with a magnification of $M \approx 22$ which results in a combined magnification of $M = 640$.

5.3. Angular momentum in the data-structures

A set of images is obtained for four typical staircase structures with different amounts of defocus. In figure 13, we show the measured data for the defocus parameter $-0.6z_r \leq z \leq 0.6z_r$, with $z_r \approx 1.9\ \text{mm}$, the Rayleigh range just in front of the lens placed on the motorized translation stage. Each column corresponds to a phase step of $\pi/2$, $3\pi/8$, $\pi/4$ and $\pi/8$ per step in the staircase structure.

To understand the measured data, a numerical comparison was done; due to the still reasonably small numerical aperture, the scalar theory is used. The interaction of the data with the structures is assumed to be adequately described by a linear phase function multiplied with

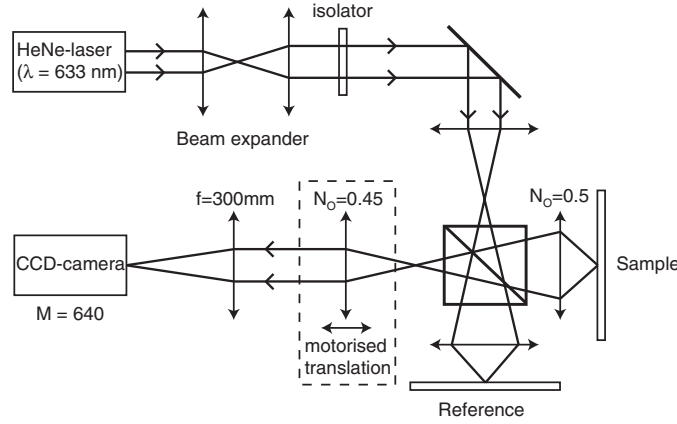


Figure 12. Schematic overview of the experimental setup. Light enters the setup emitted by a HeNe-laser, passes through beam expanding optics, followed by an optical isolator and enters the interferometer via a low numerical aperture lens. The light is focused on the sample by an $N_0 = 0.5$ microscope objective. Finally, a lens mounted on a motorized translation table re-images the light on a CCD-camera, with a magnification $M = 640$.

the illumination. The field distribution is then decomposed in Gauss–Laguerre modes due to their natural separation in cylindrical coordinates.

Starting with the paraxial situation, with a beam predominantly travelling in the positive z -direction, the paraxial wave equation can readily be derived:

$$i \frac{\partial V}{\partial z} = -\frac{1}{2\kappa} \left(\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} \right). \quad (16)$$

Next, by introducing a cylindrical coordinate system, the eigenmodes of the system of order (p, l) can be obtained as discussed in [65], yielding the well-known set of normalized Gauss–Laguerre modes:

$$V_{pl}(\mathbf{r}) = \frac{u_{pl}\sqrt{2}}{k_s w(z)} \left(\frac{2\rho^2}{w^2(z)} \right)^{|l|/2} L_p^{|l|} \left(\frac{2\rho^2}{w^2(z)} \right) \times \exp \left[il\varphi - \frac{\rho^2}{w^2(z)} - \frac{i\kappa\rho^2}{2R(z)} + i(2p + |l| + 1) \tan^{-1} \frac{z}{z_r} \right], \quad (17)$$

with

$$w(z) = \sqrt{\frac{2(z^2 + z_r^2)}{\kappa z_r}}, \quad R(z) = \frac{z^2 + z_r^2}{z}, \quad u_{pl} = (-1)^p k_s \sqrt{\frac{p!}{\pi(p + |l|)!}}, \quad (18)$$

where we have introduced $\mathbf{r} = (\rho, \varphi, z)$, the Rayleigh range z_r , the wavenumber scaled by the Rayleigh range $k_s = (\kappa/z_r)^{1/2}$ and the associated Laguerre polynomials $L_p^{|l|}(r)$.

The azimuthal index l corresponding to a contribution of $\exp[il\varphi]$, is related to the amount of orbital angular momentum in the beam. Next, only the modes with highest contributions are considered, and the modes are re-imaged for the different defocusing positions ($z = -0.6z_r$ to $z = 0.6z_r$), as shown in figure 14. The rotational sense of the beam is clearly present, and it is possible to distinguish the four different step sizes of the staircase structures. The comparison with figure 13 is relatively good, although not perfect due to always present aberrations and non-ideal depths of the individual steps. We have observed the transfer of a certain amount of angular momentum from the structures to the electromagnetic field.

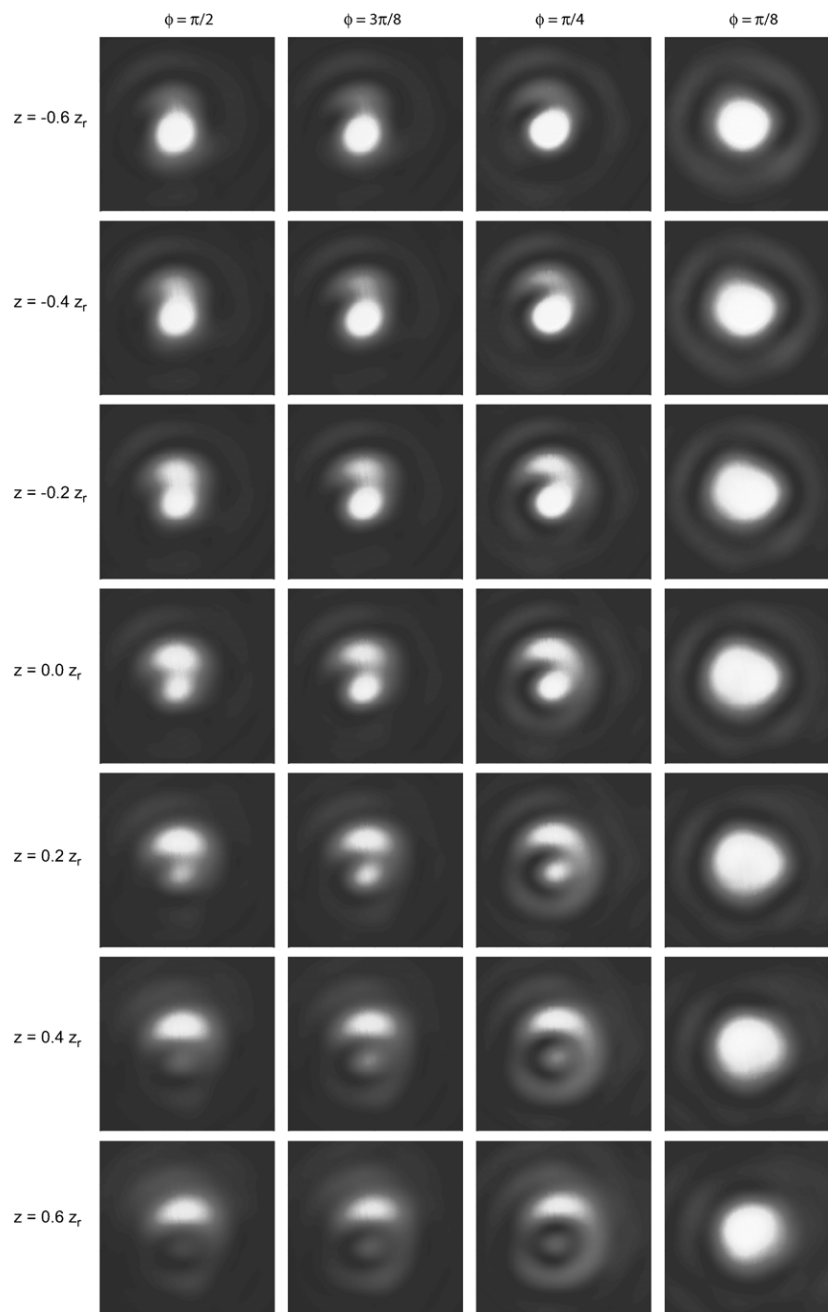


Figure 13. Experimentally obtained data for the through-focus imaging ($z = -0.6z_r$ to $z = 0.6z_r$) of four staircase configurations with $\phi = \pi/2$, $\phi = 3\pi/8$, $\phi = \pi/4$ and $\phi = \pi/8$. The defocus parameter z is normalized on the Rayleigh range $z_r \approx 1.9$ mm.

Additional insight into the field distribution after interaction with the staircase structure is gained by decomposition in the Gauss–Laguerre modes, as shown in figure 15. The mode number p is associated with the radial direction, l with the azimuthal direction. Without the staircase structure, the field would be radially symmetric, i.e. no contributions are present for

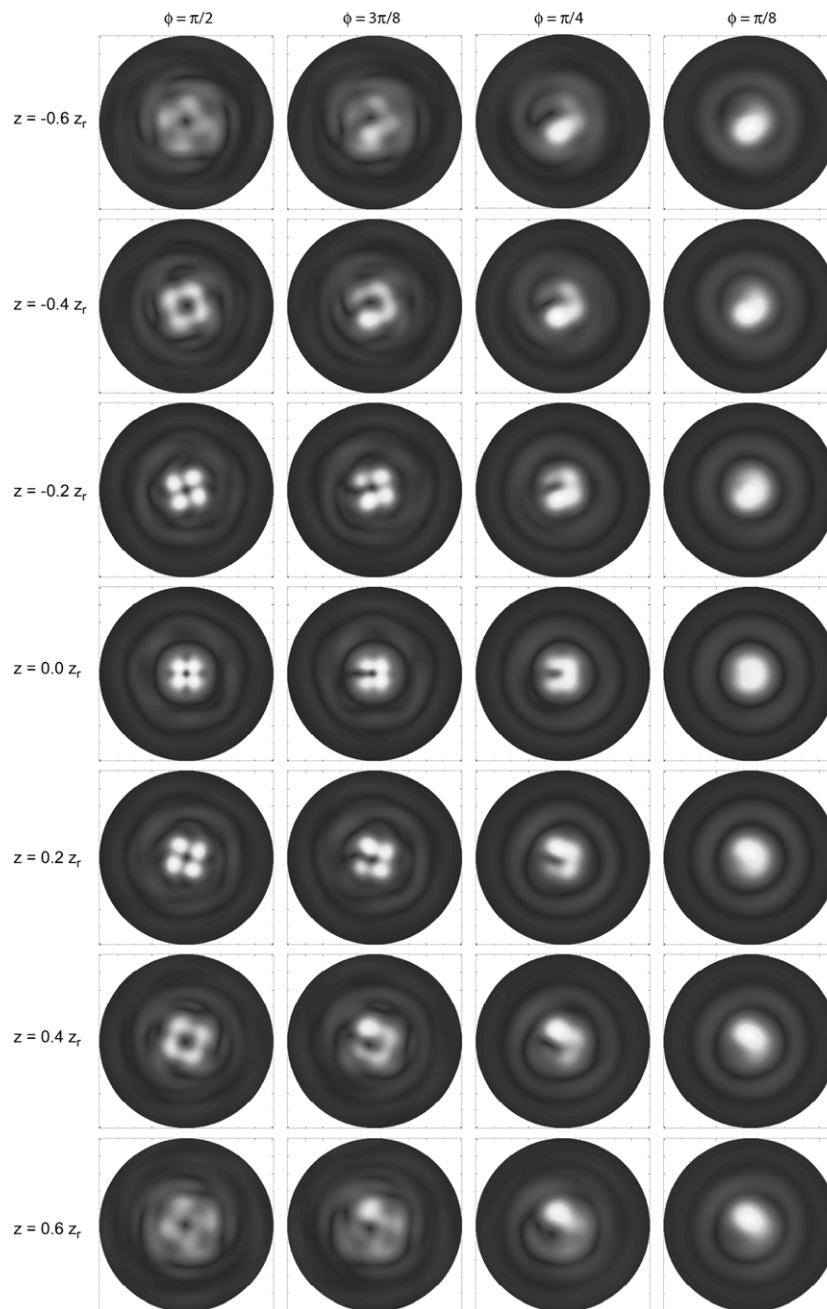


Figure 14. Simulated data for the through-focus imaging ($z = -0.6z_r$ to $z = 0.6z_r$) of four staircase configurations with $\phi = \pi/2$, $\phi = 3\pi/8$, $\phi = \pi/4$ and $\phi = \pi/8$. The defocus parameter z is normalized on the Rayleigh range $z_r \approx 1.9$ mm.

$l \neq 0$, and it is given by the point spread function according to equation (10). A decomposition in the Gauss–Laguerre modes yields a predominant contribution of $p = 0$ and smaller contributions from $p > 0$ to fit the rings around the central spot. This behaviour can be seen in figure 15(d), where small contributions are given by $l \neq 0$, indicating already an asymmetric

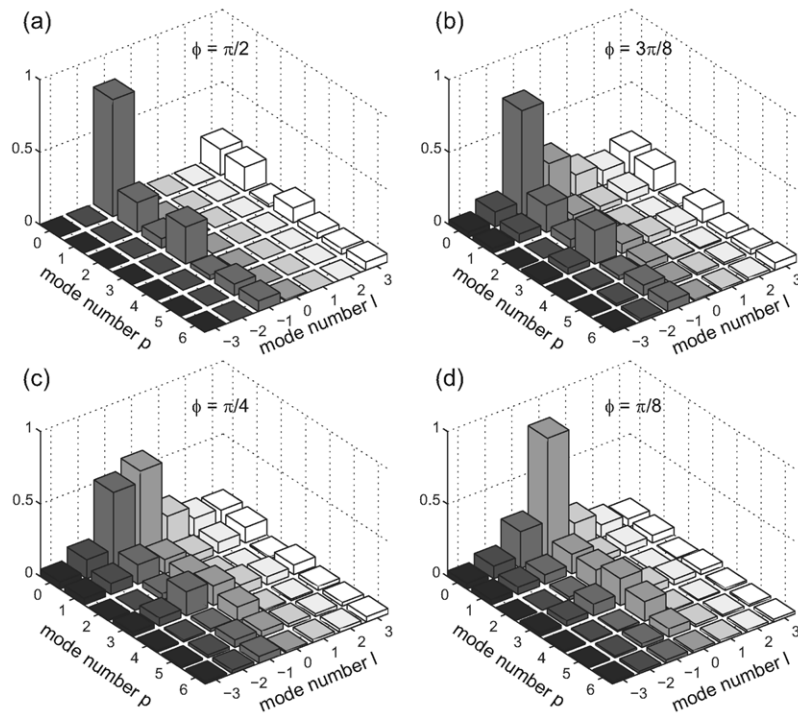


Figure 15. Modal decomposition of the light after interaction with the staircase structure with a step size of (a) $\phi = \pi/2$, (b) $\phi = 3\pi/8$, (c) $\phi = \pi/4$ and (d) $\phi = \pi/8$.

image with a certain amount of orbital angular momentum. By increasing the step size of the staircase structures, the field distribution gets more asymmetric and, for $\phi = \pi/2$, most of the light is available in the $l = -1$ modes, corresponding to the rotation of the structure. In addition to $l = -1$, there is also a small population for $l = 3$. The explanation here is that the first three steps of the staircase structure can be associated with an equal amount of gain of orbital angular momentum in the same direction; however, the last step is three times as large and corresponds to the opposite direction.

For a low numerical aperture, the measurements show good correspondence with the expected rotational behaviour of the light distribution as shown in figure 14. Although sensitive to noise by interference and residual aberrations of the imaging system, orbital angular momentum is expected to be a good candidate for optical multiplexing. Besides the possibility of changing the step size, variations in directionality and orientation of the staircase structures with respect to the track direction should be easily observable and thus can be used as independent multiplex channels. Further experiments are required to study the effect of interference by neighbouring structures.

5.4. Angular momentum in the light beam

Another candidate for optical multiplexing, based on using polarization and phase information combined in an angular momentum data-channel, moves the available intrinsic angular momentum from the structures to the illumination. Using an asymmetric structure, the reflected light can distinguish between different orientations of the structure. A promising

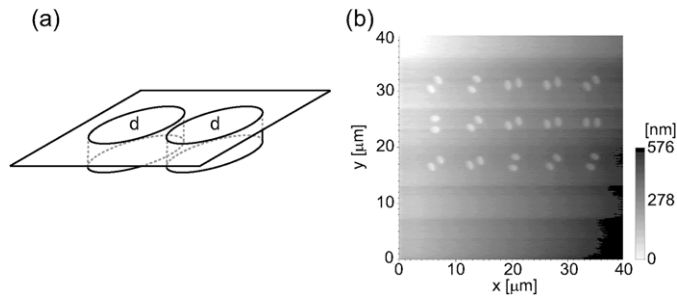


Figure 16. Test structures of dimension $2\ \mu\text{m} \times 2\ \mu\text{m}$ with two elliptically-shaped pits. (a) Schematic overview. (b) Topography of the structures obtained with an atomic force microscope.

candidate is a set of two closely spaced ellipses. Illumination of such a structure is done with a Gauss–Laguerre beam with $l \neq 0$.

A sample consisting of ellipses with a typical size of $2\ \mu\text{m} \times 2\ \mu\text{m}$ is shown in figure 16(a). These ellipses have been fabricated in the wafer with a focused ion beam. The depth is approximately 100 nm, and several ellipses are available with different orientation directions as can be seen in figure 16(b). Furthermore, it is possible to change the size, the separation and the depth of the structures to distinguish as many different information levels as possible. A Gauss–Laguerre mode with index $p = 0$ and $l = 1$ contains the required orbital angular momentum to break the symmetry of the problem. In a similar way as for the staircase structures in the previous section, the illumination is multiplied with the phase function representing the ellipsoidal structures and a decomposition in Gauss–Laguerre modes tells us the changes in the reflected field distribution. Such a mode decomposition is shown in figure 17 where the depth of the ellipses is varied from $\pi/4$ to π , corresponding to a depth change from 50 to 200 nm. With a decreasing depth of the ellipses, the modes redistribute from a range of odd l -values to the single $l = 1$ mode corresponding to the illumination. When the size of the ellipses is changed, i.e. the region where the phase difference is obtained, the field is again redistributed over the lower l -modes, but also a different modal distribution for different p -indices is observed, as shown in figure 18.

Preliminary experiments, with a Gaussian mode without intrinsic angular momentum, show that it is possible to distinguish already the different orientations of the structures over a range of π radians. Further experiments are required to study the amount of different information levels that can be distinguished, as well as the effect of noise, and the effect of neighbouring structures.

6. Conclusion

The development of optical data storage has reached a stage where further progress along the classical road of reducing the diffraction unit λ/N_O seems questionable. In this sense, a parallel between optical data storage and high-resolution lithography is evident. High-resolution lithography is the driver of Moore’s law and it has equally progressed in density through time by reducing the diffraction unit. However, because of the professional environment in which optical lithography is used, exotic wavelengths like 193 nm (KrF-laser) or even 13.5 nm (plasma-generated extreme UV) are possible. Liquid immersion has also been successfully applied in optical lithography to further increase the capacity of integrated circuits. In optical data storage, the application of wavelengths beyond the blue region seems questionable, among

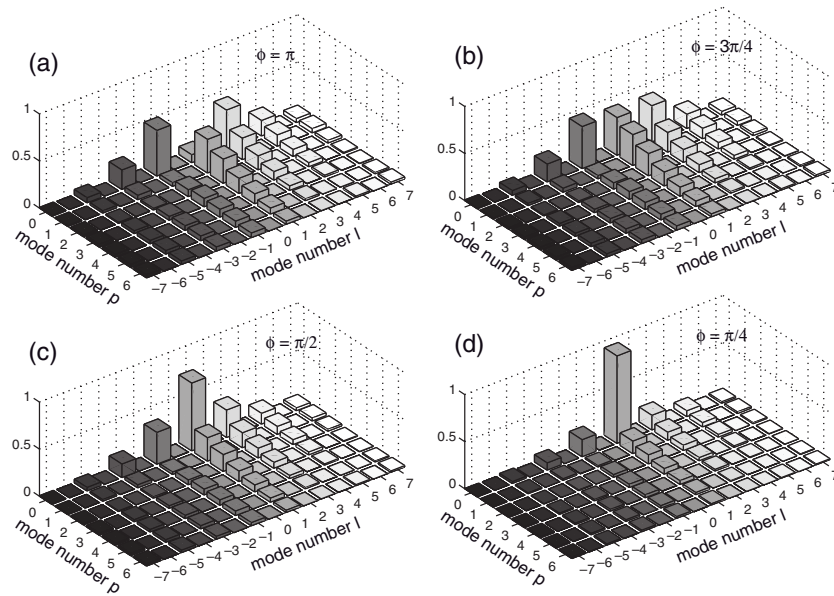


Figure 17. Decomposition in Gauss-Laguerre modes of the ellipses multiplied with the incident illumination distribution. The depth of the structure is varied, the size of the ellipses is kept constant at $\omega = \omega_0$.

other reasons because of the limited lifetime of plastic and other amorphous glass-based optical materials when exposed to light in the deep UV spectral region. The increase of numerical aperture beyond unity is possible using the principle of frustrated internal reflection by making the light tunnel through a very thin air gap between the last surface of the objective and the data- or cover-layer of the optical disc. The air gap is typically less than 50 nm and the optical information is read-out in the so-called ‘solid’ immersion. Such a high-density system has to cope with extra boundary conditions (ultra-flat disc surface, dust protection, contamination issues) that are not easily solved for a mass-consumer product. But preliminary experiments on the solid-immersion option show that this approach might be more realistic for widespread application than was initially thought. Even the intrinsically low modulation depth of the detected signal is not a basic problem in the presence of modern detection techniques. Other prototype high-density data storage systems are based on a holographic storage technique. Early experiments at the beginning of the 1970s already used the holographic principle, albeit in a rather primitive way. For example, pictures like TV-frames were directly recorded onto the holographic medium without any picture coding. Nowadays techniques use coded bit patterns to be stored holographically so that cross-talk between images can be repaired in the digital domain using decoding and error correction techniques. But the basic drawback of a holographic storage medium remains its complexity. The system needs a 2D pixelized light modulator for recording and a 2D-detector for the reconstruction. Moreover, the proposed holographic systems rely on angular multiplexing (beam deflection) for superimposing the holograms in the recording medium and this introduces yet another expensive optical element. For these reasons, it seems highly improbable that the actual holographic systems have a serious chance of penetrating the mass-consumer storage market.

In this paper, we have discussed alternative methods for obtaining high density discs, using the multiplex principle. In our proposal we have introduced a fine structure in the recorded optical effects so that more bits than one can be stored per location. A specific way

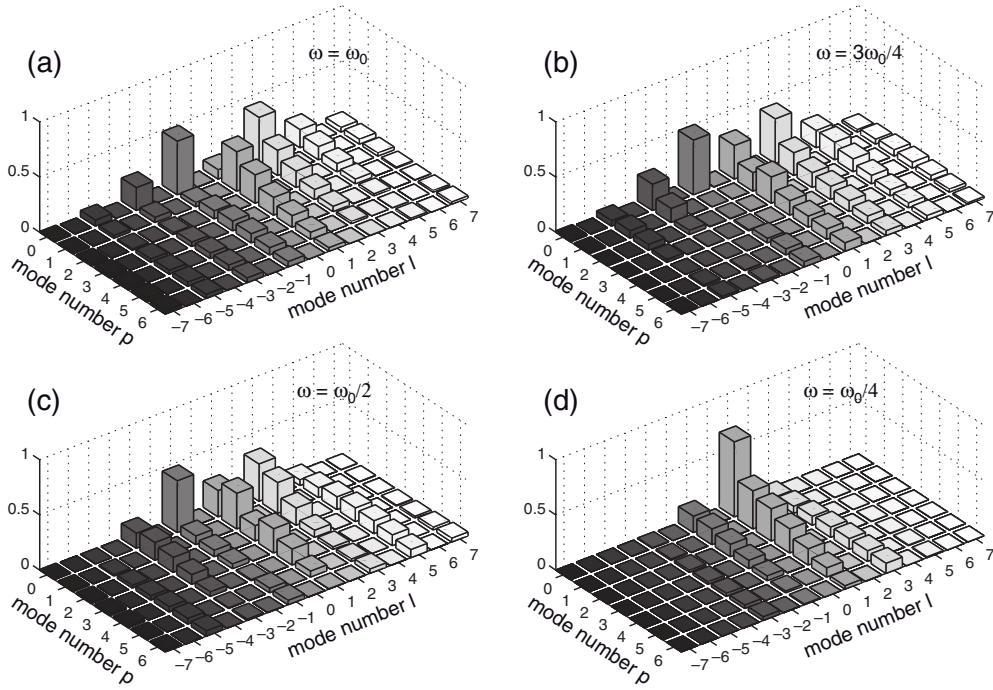


Figure 18. Decomposition in Gauss–Laguerre modes of the ellipses multiplied with the incident illumination distribution. The depth of the structure is kept constant at $\phi = \pi$, the size of the ellipses is varied.

of multiplexing was based on the introduction of angular momentum in the reflected beam by making the optical effects impart a helical phase pattern on reflection. One extra bit is possible by choosing the screw sense of the phase as an extra degree of freedom. Another degree of freedom is the azimuth of the specially designed optical effects. In this way, a substantial increase in storage capacity should be feasible. As an extra advantage of the multiplexing approach we mention the automatic increase of the data retrieval rate. In the actual high-density systems, data retrieval rate is becoming a problem and the multiplexing option offers an interesting relief in this respect.

Throughout our analysis of the high-density system, we have employed rigorous methods to describe the electromagnetic field in focus in the multilayered optical disc structure. We have also used a rigorous description of the diffraction process at the sub-wavelength structures which contain the stored information. Our preferred tool has been a three-dimensional electromagnetic version of Green's tensor formalism. It is shown by means of calculated read-out signals that, in high-density optical systems, the commonly employed scalar diffraction theory using Kirchhoff boundary conditions is inadequate and that vector diffraction methods need to be applied.

Appendix A. Field in the focal region

This appendix provides an explicit expression for the transmission and reflection matrices required to calculate the field in the focal region in stratified media, as described in [39, 66] and a treatment of the Strehl intensity in the high numerical aperture case.

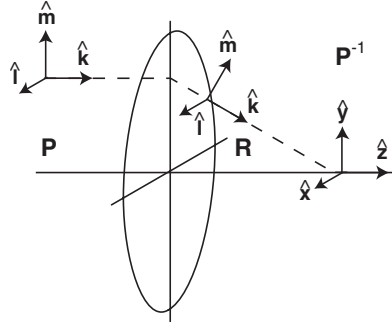


Figure A1. Schematic overview of the projection operator \mathbf{P} and the rotation operator \mathbf{R} . The Cartesian basis $(\hat{x}, \hat{y}, \hat{z})$ and natural basis $(\hat{k}, \hat{l}, \hat{m})$ have been introduced to describe the lens operator $\mathbf{L} = \mathbf{P}^{-1}\mathbf{R}\mathbf{P}$.

Appendix A.1. Lens rotation matrix

To be able to include the rotation effect of the lens on the field vector, we project the Cartesian basis $(\hat{x}, \hat{y}, \hat{z})$ on a cylindrical basis $(\hat{k}, \hat{l}, \hat{m})$ denoted by the operation \mathbf{P} , perform a rotation of the field vectors in the cylindrical basis denoted by the operation \mathbf{R} and project the cylindrical basis back onto the Cartesian basis; this latter operation is the inverse of the first operation \mathbf{P}^{-1} , yielding a lens operator matrix \mathbf{L} , as depicted in figure A1. The matrices describing the operations \mathbf{P} , \mathbf{R} and \mathbf{L} are given by

$$\mathbf{P} = \begin{pmatrix} \cos k_\phi & \sin k_\phi & 0 \\ -\sin k_\phi & \cos k_\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (\text{A.1})$$

$$\mathbf{R} = \frac{1}{k_1} \begin{pmatrix} k_{z1} & 0 & -k_r \\ 0 & k_1 & 0 \\ k_r & 0 & k_{z1} \end{pmatrix}, \quad (\text{A.2})$$

$$\mathbf{L} = \mathbf{P}^{-1}\mathbf{R}\mathbf{P} = \frac{1}{k_1} \begin{pmatrix} k_{z1} \cos^2 k_\phi + k_1 \sin^2 k_\phi & (k_{z1} - k_1) \sin k_\phi \cos k_\phi & -k_r \cos k_\phi \\ (k_{z1} - k_1) \sin k_\phi \cos k_\phi & k_{z1} \sin^2 k_\phi + k_1 \cos^2 k_\phi & -k_r \sin k_\phi \\ k_r \cos k_\phi & k_r \sin k_\phi & k_{z1} \end{pmatrix}, \quad (\text{A.3})$$

where we have introduced the propagation vector $\mathbf{k} = (k_r, k_\phi, \pm k_z)$ in cylindrical coordinates and identify $k_1 = |\mathbf{k}_1|$ as the modulus of the propagation vector in the first medium. It remains to propagate the field through the several planar medium transitions with their surface normals pointing in the z -direction.

Appendix A.2. Layer transitions

Therefore, we start by studying an arbitrary plane wave with wave vector \mathbf{k}_i

$$\mathbf{E}(\mathbf{k}_i^\pm) = \mathbf{A} e^{\pm i \mathbf{k}_i^\pm \cdot \mathbf{r}}. \quad (\text{A.4})$$

The subscript i from the propagation vector $\mathbf{k}_i^\pm = (k_{ri}, k_{\phi i}, \pm k_{zi})$ denotes the medium in the medium in which the vector is defined and the superscript \pm denotes the sign of the propagation direction. Since the transition of one medium i to the next medium j at the plane $z = d_{i,j}$ should be independent of the spatial cylindrical coordinates $\mathbf{r} = (\rho, \varphi, d_{i,j})$, it follows from equation (A.4) that $k_{ri} = k_{rj} = k_r$ and $k_{\phi i} = k_{\phi j} = k_\phi$. For monochromatic waves, as is the

case considered here, we have a fixed length of the propagation vector $k_i^2 = k_{zi}^2 + k_r^2$. The sign of the square root which should be taken to acquire the propagation vector k_{zi} in the z -direction follows from energy conservation; i.e. the sign should be chosen such that the wave is exponentially decreasing in the direction of propagation. In the cylindrical basis, the unit vector $\hat{\mathbf{k}} = \mathbf{k}/k$ is parallel to the direction of propagation of the wave, the unit vector $\hat{\mathbf{l}} = (\hat{\mathbf{k}} \times \hat{\mathbf{z}})/|\hat{\mathbf{k}} \times \hat{\mathbf{z}}|$ is perpendicular to the propagation vector and the normal to the interface (TE-polarization) and the unit vector $\hat{\mathbf{m}} = \hat{\mathbf{k}} \times \hat{\mathbf{l}}$ completes the orthonormal basis (TM-polarization).

$$\hat{\mathbf{k}}_i^\pm = \frac{1}{k_i} \begin{pmatrix} k_r \cos k_\phi \\ k_r \sin k_\phi \\ \pm k_{zi} \end{pmatrix}, \quad \hat{\mathbf{l}}_i^\pm = \hat{\mathbf{l}} = \begin{pmatrix} \sin k_\phi \\ -\cos k_\phi \\ 0 \end{pmatrix}, \quad \hat{\mathbf{m}}_i^\pm = \frac{1}{k_i} \begin{pmatrix} \pm k_{zi} \cos k_\phi \\ \pm k_{zi} \sin k_\phi \\ -k_r \end{pmatrix}. \quad (\text{A.5})$$

By performing a scalar multiplication of the field vector with the three unit vectors as defined in equation (A.5), we obtain three components associated with the oscillation direction.

For the transition from medium i to the next medium j , we multiply these components with the corresponding unit vectors; performing both operations at once yields three matrices $\hat{\mathbf{k}}_j^\pm \hat{\mathbf{k}}_i^\pm$, $\mathbf{l}_j \mathbf{l}_i$ and $\hat{\mathbf{m}}_j^\pm \hat{\mathbf{m}}_i^\pm$, resulting in

$$\hat{\mathbf{k}}_j^\pm \hat{\mathbf{k}}_i^\pm = \frac{1}{k_i k_j} \begin{pmatrix} k_r^2 \cos^2 k_\phi & k_r^2 \sin k_\phi \cos k_\phi & k_r k_{zi} \cos k_\phi \\ k_r^2 \sin k_\phi \cos k_\phi & k_r^2 \sin^2 k_\phi & k_r k_{zi} \sin k_\phi \\ \pm k_r k_{zj} \cos k_\phi & \pm k_r k_{zj} \sin k_\phi & \pm k_{zi} k_{zj} \end{pmatrix}, \quad (\text{A.6a})$$

$$\hat{\mathbf{u}} = \begin{pmatrix} \sin^2 k_\phi & -\sin k_\phi \cos k_\phi & 0 \\ -\sin k_\phi \cos k_\phi & \cos^2 k_\phi & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (\text{A.6b})$$

$$\hat{\mathbf{m}}_j^\pm \hat{\mathbf{m}}_i^\pm = \frac{1}{k_i k_j} \begin{pmatrix} \pm k_{zi} k_{zj} \cos^2 k_\phi & \pm k_{zi} k_{zj} \sin k_\phi \cos k_\phi & \mp k_r k_{zj} \cos k_\phi \\ \pm k_{zi} k_{zj} \sin k_\phi \cos k_\phi & \pm k_{zi} k_{zj} \sin^2 k_\phi & \mp k_r k_{zj} \sin k_\phi \\ -k_r k_{zi} \cos k_\phi & -k_r k_{zi} \sin k_\phi & k_r^2 \end{pmatrix}. \quad (\text{A.6c})$$

The unit vector $\hat{\mathbf{l}}$ is independent of both the medium and the direction of propagation, therefore the same holds for matrix $\hat{\mathbf{u}}$. We know that $\hat{\mathbf{k}}_i^\pm \hat{\mathbf{k}}_i^\pm + \hat{\mathbf{l}}_i^\pm \hat{\mathbf{l}}_i^\pm + \hat{\mathbf{m}}_i^\pm \hat{\mathbf{m}}_i^\pm = \mathbf{I}$, the identity matrix.

Next, we have to derive the transmission and reflection coefficients for optical waves. The Maxwell equations yielding the boundary conditions between medium i and medium j are given by

$$\epsilon_i E_{\perp i} = \epsilon_j E_{\perp j}, \quad E_{\parallel i} = E_{\parallel j}, \quad \mu_i H_{\perp i} = \mu_j H_{\perp j}, \quad H_{\parallel i} = H_{\parallel j}. \quad (\text{A.7})$$

We obtain two independent equations for the transverse electric (TE) and for the transverse magnetic (TM) polarized components. For the TE-polarization (s), the electric and magnetic field components read as

$$E_{\perp i} = 0, \quad E_{\parallel i} = E_i^s, \quad H_{\perp i} = \frac{k_r}{\omega \mu_0} E_i^s, \quad H_{\parallel i} = -\frac{k_{zi}}{\omega \mu_0} E_i^s, \quad (\text{A.8})$$

where we have used $i\omega\mu_0\mathbf{H}_i = i\mathbf{k}_i \times \mathbf{E}_i$ from the Maxwell arbitrary plane wave as defined by equation (A.4) with $\mu = \mu_0$. Combined with the boundary conditions, we obtain the relations

$$A_i^{s+} e^{ik_{zi}d_{i,j}} + A_i^{s-} e^{-ik_{zi}d_{i,j}} = A_j^{s+} e^{ik_{zj}d_{i,j}} + A_j^{s-} e^{-ik_{zj}d_{i,j}}, \quad (\text{A.9a})$$

$$k_{zi} A_i^{s+} e^{ik_{zi}d_{i,j}} - k_{zi} A_i^{s-} e^{-ik_{zi}d_{i,j}} = k_{zj} A_j^{s+} e^{ik_{zj}d_{i,j}} - k_{zj} A_j^{s-} e^{-ik_{zj}d_{i,j}}, \quad (\text{A.9b})$$

and for the TM polarization (p) the electric and magnetic field components read as

$$E_{\perp i} = -\frac{k_r}{k_i} E_i^p, \quad E_{\parallel i} = \frac{k_{zi}}{k_i} E_i^p, \quad H_{\perp i} = 0, \quad H_{\parallel i} = \frac{k_i}{\omega \mu_0} E_i^p, \quad (\text{A.10})$$

which yields the relations

$$\frac{\epsilon_i}{k_i} A_i^{p+} e^{ik_{zi}d_{i,j}} + \frac{\epsilon_i}{k_i} A_i^{p-} e^{-ik_{zi}d_{i,j}} = \frac{\epsilon_j}{k_j} A_j^{p+} e^{ik_{zj}d_{i,j}} + \frac{\epsilon_j}{k_j} A_j^{p-} e^{-ik_{zj}d_{i,j}}, \quad (\text{A.11a})$$

$$\frac{k_{zi}}{k_i} A_i^{p+} e^{ik_{zi}d_{i,j}} - \frac{k_{zi}}{k_i} A_i^{p-} e^{-ik_{zi}d_{i,j}} = \frac{k_{zj}}{k_j} A_j^{p+} e^{ik_{zj}d_{i,j}} - \frac{k_{zj}}{k_j} A_j^{p-} e^{-ik_{zj}d_{i,j}}. \quad (\text{A.11b})$$

Solving both sets of equations for $A_i^{s/p-}$ and $A_i^{s/p+}$, we obtain the ratio of backward travelling waves over forward travelling waves:

$$(A^\pm/A^\mp)_i^{s/p} = \frac{F_{i,j}^{s/p} e^{\mp ik_{zj}d_{i,j}} + (A^\pm/A^\mp)_j^{s/p} e^{\pm ik_{zj}d_{i,j}}}{e^{\mp ik_{zj}d_{i,j}} + (A^\pm/A^\mp)_j^{s/p} F_{i,j}^{s/p} e^{\pm ik_{zj}d_{i,j}}}. \quad (\text{A.12})$$

The Fresnel reflection coefficients for the TE and TM polarizations are given by

$$F_{i,j}^s = \frac{k_{zi} - k_{zj}}{k_{zi} + k_{zj}}, \quad (\text{A.13a})$$

$$F_{i,j}^p = \frac{\epsilon_j k_{zi} - \epsilon_i k_{zj}}{\epsilon_j k_{zi} + \epsilon_i k_{zj}}. \quad (\text{A.13b})$$

Since we know that in the last medium with $i = N$ there are no backward travelling waves coming from infinity, the ratio $(A^-/A^+)_N^{s/p} = 0$. In the first medium with $i = 1$, the same holds for the forward over backward travelling waves $(A^+/A^-)_1^{s/p} = 0$.

Next, we propagate the field calculated in layer i to the next layer j . The effective reflection and transmission coefficients pertaining to a general transition (i, j) can be calculated by progression through the successive layers using equations (A.9a) and (A.11a):

$$A_j^{s/p\pm} = \gamma_{i,j}^{s/p} \left[A_i^{s/p\pm} e^{\pm ik_{zi}d_{i,j}} - A_i^{s/p\mp} F_{i,j}^{s/p} e^{\mp ik_{zi}d_{i,j}} \right] e^{\mp ik_{zj}d_{i,j}}. \quad (\text{A.14})$$

The polarization-dependent pre-factor is defined as

$$\gamma_{i,j}^s = \frac{k_{zi} + k_{zj}}{2k_{zj}}, \quad (\text{A.15a})$$

$$\gamma_{i,j}^p = \frac{k_j \epsilon_j k_{zi} + \epsilon_i k_{zj}}{k_i 2\epsilon_j k_{zj}}. \quad (\text{A.15b})$$

The progression has to be continued until the final layer where the observation point \mathbf{r} is located. Note that for p -polarization each layer transition is accompanied by a change of the basis vectors, which effect is incorporated in the matrix multiplication:

$$\hat{\mathbf{m}}_j^\pm \hat{\mathbf{m}}_i^\pm = \hat{\mathbf{m}}_j^\pm \hat{\mathbf{m}}_{j-1}^\pm \cdots \hat{\mathbf{m}}_{i+2}^\pm \hat{\mathbf{m}}_{i+1}^\pm \cdot \hat{\mathbf{m}}_{i+1}^\pm \hat{\mathbf{m}}_i^\pm. \quad (\text{A.16})$$

The k_ϕ dependence in the projection matrices in equations (B.17a) and (B.17b), using

$$\int_0^{2\pi} e^{imk_\phi} e^{i\rho k_r \cos(k_\phi - \phi)} dk_\phi = 2\pi i^m J_m(\rho k_r) e^{im\phi}, \quad (\text{A.17})$$

can be integrated analytically:

$$\int_0^{2\pi} \hat{\mathbf{u}} e^{i\rho k_r \cos(k_\phi - \phi)} dk_\phi = \frac{2\pi}{2} \begin{pmatrix} J_0 + J_2 \cos 2\varphi & J_2 \sin 2\varphi & 0 \\ J_2 \sin 2\varphi & J_0 - J_2 \cos 2\varphi & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (\text{A.18})$$

$$\begin{aligned} & \int_0^{2\pi} \hat{\mathbf{m}}_j^\pm \hat{\mathbf{m}}_i^\pm e^{i\rho k_r \cos(k_\phi - \phi)} dk_\phi \\ &= \frac{2\pi}{k_i k_j} \begin{pmatrix} \pm \frac{k_{zi} k_{zj}}{2} (J_0 - J_2 \cos 2\varphi) & \mp \frac{k_{zi} k_{zj}}{2} J_2 \sin 2\varphi & \mp i k_r k_{zj} J_1 \cos \varphi \\ \mp \frac{k_{zi} k_{zj}}{2} J_2 \sin 2\varphi & \pm \frac{k_{zi} k_{zj}}{2} (J_0 + J_2 \cos 2\varphi) & \mp i k_r k_{zj} J_1 \sin \varphi \\ -i k_r k_{zi} J_1 \cos \varphi & -i k_r k_{zi} J_1 \sin \varphi & k_r^2 J_0 \end{pmatrix}, \quad (\text{A.19}) \end{aligned}$$

where, for clarity, the argument of the Bessel functions $J_m(\rho k_r)$ has been omitted. The k_r dependence has to be integrated numerically. We have used an adaptive Gauss–Kronrod quadrature for this purpose.

Appendix A.3. Generalized lens matrix

In general, the lens operator could include several layer transitions to be able to calculate the field in layer i . Using equations (A.6a)–(A.6c), we obtain

$$\mathbf{M}_i^{z\pm} = f_i^{z\pm} \cdot \hat{k}_i^\pm \hat{k}_1 \cdot \mathbf{L} = \frac{f_i^{z\pm}}{k_i} \begin{pmatrix} 0 & 0 & -k_r \cos k_\phi \\ 0 & 0 & -k_r \sin k_\phi \\ 0 & 0 & \pm k_{zi} \end{pmatrix}, \quad (\text{A.20a})$$

$$\mathbf{M}_i^{s\pm} = f_i^{s\pm} \cdot \hat{l} \cdot \mathbf{L} = f_i^{s\pm} \begin{pmatrix} \sin^2 k_\phi & -\sin k_\phi \cos k_\phi & 0 \\ -\sin k_\phi \cos k_\phi & \cos^2 k_\phi & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (\text{A.20b})$$

$$\mathbf{M}_i^{p\pm} = f_i^{p\pm} \cdot \hat{m}_i^\pm \hat{m}_1 \cdot \mathbf{L} = \frac{f_i^{p\pm}}{k_i} \begin{pmatrix} \pm k_{zi} \cos^2 k_\phi & \pm k_{zi} \cos k_\phi \sin k_\phi & 0 \\ \pm k_{zi} \cos k_\phi \sin k_\phi & \pm k_{zi} \sin^2 k_\phi & 0 \\ k_r \cos k_\phi & k_r \sin k_\phi & 0 \end{pmatrix}, \quad (\text{A.20c})$$

where $f_i^{z\pm}$, $f_i^{s\pm}$ and $f_i^{p\pm}$ denote the effective transmission and reflection coefficients for the longitudinal, TE and TM polarization components, respectively. Since optical waves travelling in the longitudinal direction are purely transversal, we can set the effective transmission and reflection coefficients for the longitudinal contribution to zero, $f_i^{z\pm} = 0$, therefore the contribution of the matrix $\mathbf{M}_i^{z\pm} = 0$.

Next, we define $f_i^{s/p\pm} = A_i^{s/p\pm}/A_1^{s/p\pm}$, yielding for the effective transmission coefficient in the first medium $f_1^{s/p+} = 1$ and for effective reflection coefficient $f_1^{s/p-} = (A^-/A^+)_1$. The fraction $(A^-/A^+)_1$ is obtained by using equation (A.12) in an iterative procedure, starting in layer N with $(A^-/A^+)_N = 0$. Now we rewrite equation (A.14) as

$$f_{i+1}^{s/p\pm} = \gamma_{i,i+1}^{s/p} \left[f_i^{s/p\pm} e^{\pm i k_{zi} d_{i,i+1}} - f_i^{s/p\mp} F_{i,i+1}^{s/p} e^{\mp i k_{zi} d_{i,i+1}} \right] e^{\mp i k_{zi+1} d_{i,i+1}}. \quad (\text{A.21})$$

Finally, we find for the total propagation matrix

$$\mathbf{M}_i^\pm = \mathbf{M}_i^{p\pm} + \mathbf{M}_i^{s\pm} = \begin{pmatrix} g_i^{0\pm} - g_i^{2\pm} \cos 2k_\phi & -g_i^{2\pm} \sin 2k_\phi & 0 \\ -g_i^{2\pm} \sin 2k_\phi & g_i^{0\pm} + g_i^{2\pm} \cos 2k_\phi & 0 \\ g_i^{1\pm} \cos k_\phi & g_i^{1\pm} \sin k_\phi & 0 \end{pmatrix}, \quad (\text{A.22})$$

with the generalized reflection and transmission coefficients given by

$$g_i^{0\pm} = \frac{1}{2} \left(f_i^{s\pm} \pm f_i^{p\pm} \frac{k_{zi}}{k_i} \right), \quad (\text{A.23a})$$

$$g_i^{2\pm} = \frac{1}{2} \left(f_i^{s\pm} \mp f_i^{p\pm} \frac{k_{zi}}{k_i} \right), \quad (\text{A.23b})$$

$$g_i^{1\pm} = f_i^{p\pm} \frac{k_r}{k_i}, \quad (\text{A.23c})$$

where the number $n = 0, 1, 2$ relates to the angular dependence of the sine- or cosine-function that the coefficient is multiplied with in the propagation matrices for the integration over k_ϕ . Note that for a single medium $N = 1$, $\mathbf{M}_1^+ = \mathbf{M}$ and $\mathbf{M}_1^- = 0$, where the lens rotation matrix \mathbf{L}

would be equal to the propagation matrix \mathbf{M} if the last column is taken equal to zero. Finally, since $\mathbf{a}(k_r, k_\phi) = \mathbf{M}(k_r, k_\phi + \pi) \cdot \mathbf{E}(k_r, k_\phi)$, the general propagation matrix is given by

$$\mathbf{M}_i^\pm(k_r, k_\phi) = \begin{pmatrix} g_i^{0\pm} - g_i^{2\pm} \cos 2k_\phi & -g_i^{2\pm} \sin 2k_\phi & 0 \\ -g_i^{2\pm} \sin 2k_\phi & g_i^{0\pm} + g_i^{2\pm} \cos 2k_\phi & 0 \\ -g_i^{1\pm} \cos k_\phi & -g_i^{1\pm} \sin k_\phi & 0 \end{pmatrix}. \quad (\text{A.24})$$

Appendix A.4. The definition and calculation of Strehl intensity in the high-numerical-aperture case

In the scalar case, applicable to numerical apertures as small as 0.40 to 0.50, the axial or Strehl intensity is defined by

$$I_S(0, 0) = \frac{\left| \frac{1}{\pi} \int \int_{S_{\text{pupil}}} \exp[i f \rho^2 + i \Phi(\rho, \theta)] \rho d\rho d\theta \right|^2}{\frac{1}{\pi} \int \int_{S_{\text{pupil}}} \rho d\rho d\theta}. \quad (\text{A.25})$$

In this expression we have used polar coordinates (ρ, θ) on the pupil sphere with ρ normalized to half the diameter $2\rho_{\text{max}}$ of the exit pupil. The term with the defocus parameter f allows an adjustment of the focus setting by an amount Δz with respect to the nominal position at recording or read-out in order to optimize the Strehl intensity; f is given $\pi z N_0^2 / \lambda$. The aberration function Φ accounts for the phase aberration introduced by the optical system. For a perfect optical imaging system in optimum focus ($f = 0$), the value of I_S amounts to unity.

In expression (A.25) we have supposed a uniform intensity distribution over the exit pupil. An extension to the more general situation of varying amplitude strength over the pupil cross-section leads to

$$I_S(0, 0) = \frac{\left| \frac{1}{\pi} \int \int_{S_{\text{pupil}}} A(\rho, \theta) \exp[i f \rho^2 + i \Phi(\rho, \theta)] \rho d\rho d\theta \right|^2}{\frac{1}{\pi} \int \int_{S_{\text{pupil}}} |A(\rho, \theta)|^2 \rho d\rho d\theta}. \quad (\text{A.26})$$

The definition of Strehl intensity is preferably limited to small values of Φ , typically of the order of unity or less. In that case we can expand the exponential factor in equation (A.26) up to second order and the average complex amplitude $A_S(0, 0)$ between the modulus signs is given by

$$\begin{aligned} A_S(0, 0) &= \frac{1}{\pi} \int \int_{S_{\text{pupil}}} A(\rho, \theta) \exp[i f \rho^2 + i \Phi(\rho, \theta)] \rho d\rho d\theta \\ &= \frac{1}{\pi} \int \int_{S_{\text{pupil}}} A(\rho, \theta) \left[1 + i(f\rho^2 + \Phi) - \frac{1}{2}(f\rho^2 + \Phi)^2 \right] \rho d\rho d\theta \\ &= \left[\overline{A} - \frac{1}{2} \overline{A\Phi_f^2} + i \overline{A\Phi_f} \right], \end{aligned} \quad (\text{A.27})$$

where we have first substituted $\Phi_f(\rho, \theta) = \Phi(\rho, \theta) - f\rho^2$ and then omitted the coordinate-dependence in the expression.

The expression for the Strehl intensity can now be written as

$$I_S(0, 0) = \frac{(\overline{A})^2 - \overline{A\Phi_f^2} + (\overline{A\Phi_f})^2}{\overline{A^2}}. \quad (\text{A.28})$$

An optimization of the Strehl intensity can be carried out by adapting the defocus parameter f . Even for a perfectly corrected optical system, $I_S(0, 0)$ cannot exceed the value $(\overline{A})^2 / \overline{A^2}$.

Appendix A.4.1. Strehl intensity for a high-numerical-aperture focused beam. The extension of the definition of Strehl intensity to the high-numerical-aperture case needs the following adaptations due to the vector diffraction analysis:

- the Strehl intensity is proportional to the modulus squared of the electric field vector;

$$I_S(0, 0) \propto |\mathbf{E}|^2; \quad (\text{A.29})$$

here, we can use the fact that the electric field components that are not parallel to the incident linear polarization vector are all zero in the on-axis focal point.

- the contribution to the non-zero electric field component in the axial focal point is not constant over the exit pupil cross-section; it is determined by a factor A_p that in the case of x -polarized incident light is given by

$$A_p(\rho, \theta) = \frac{1}{2} \left(1 + \sqrt{1 - N_O^2 \rho^2} \right) - \frac{1}{2} \left(1 - \sqrt{1 - N_O^2 \rho^2} \right) \cos 2\theta. \quad (\text{A.30})$$

- the mapping of the amplitude of a uniform plane wave in the entrance pupil onto the exit pupil leads to an amplitude non-uniformity given by $(1 - N_O \rho^2)^{-1/4}$, the so-called radiometric effect (see [29, 30]).

The combined amplitude function including radiometric effect and a linear incident state of polarization (x -direction) reads as

$$A_x(\rho, \theta) = \frac{\left(1 + \sqrt{1 - N_O^2 \rho^2} \right) - \left(1 - \sqrt{1 - N_O^2 \rho^2} \right) \cos 2\theta}{2(1 - N_O^2 \rho^2)^{1/4}}. \quad (\text{A.31})$$

In the case of an aberration-free optical system we find the following expression for the Strehl-intensity [67]

$$I_S(0, 0) = \left(\frac{8}{75N_O^2} \right) \frac{[8 - 5(1 - N_O^2)^{3/4} - 3(1 - N_O^2)^{5/4}]^2}{4 + 3N_O^2 - (1 - N_O^2)^{1/2}(4 - N_O^2)}. \quad (\text{A.32})$$

As an example we consider the value of $N_O = 0.95$. The corresponding on-axis intensity is 1.05856 (numerator of equation (A.32)) and, after normalization to the incident power (1.060075), we find a Strehl intensity I_S equal to 0.99857.

A further deviation from scalar theory is the exact description of the defocusing phase factor at high numerical aperture. The phase aberration due to a defocusing over a distance z is exactly given by

$$\Phi_f = z \frac{2\pi}{\lambda} \left(1 - \sqrt{1 - N_O^2 \rho^2} \right), \quad (\text{A.33})$$

showing that the defocusing term is not simply described by a quadratic phase factor. At high values of N_O , substantial values of order four and higher need to be incorporated in the series expansion of the square root expression.

Appendix A.4.2. Strehl intensity of a generally polarized incident beam. We write the incident electric field as $\mathbf{E} = (E_x, E_y, 0)$ and, with reference to equation (A.27), we find for the electric

field on axis

$$\begin{aligned} \mathbf{E}_S(0, 0) = & \left[\overline{E_x A_x} - \frac{1}{2} \overline{E_x A_x \Phi_f^2} + i \overline{E_x A_x \Phi_f} \right] \hat{\mathbf{e}}_x \\ & + \left[\overline{E_y A_y} - \frac{1}{2} \overline{E_y A_y \Phi_f^2} + i \overline{E_y A_y \Phi_f} \right] \hat{\mathbf{e}}_y, \end{aligned} \quad (\text{A.34})$$

where $\hat{\mathbf{e}}_x$ and $\hat{\mathbf{e}}_y$ are the unit vectors in the x - and y -directions and A_y is obtained from equation (A.31) by replacing θ by $(\theta - \pi/2)$. In the general case, E_x and E_y are functions of (ρ, θ) and can have a varying phase difference, allowing a very general state of polarization of the wave in the exit pupil or on the focusing region [68].

The Strehl intensity is obtained by evaluating the modulus squared $\mathbf{E}_S(0, 0)$ and this yields

$$\begin{aligned} I_S(0, 0) = & \left[(\overline{E_x A_x})^2 + (\overline{E_y A_y})^2 - \overline{E_x A_x \Phi_f^2} - \overline{E_y A_y \Phi_f^2} \right. \\ & \left. + (\overline{E_x A_x \Phi_f})^2 + (\overline{E_y A_y \Phi_f})^2 \right] / \left[(\overline{E_x A_x})^2 + (\overline{E_y A_y})^2 \right]. \end{aligned} \quad (\text{A.35})$$

We conclude from equation (A.35) that the contributions from orthogonal polarization components in the entrance pupil add incoherently to the Strehl intensity in the image plane. The exact optimization of Strehl intensity for optimum read-out is outside the scope of this paper. The foregoing analysis is meant to give the framework within which wavefront minimization has to be carried out.

Appendix B. Solution to the general scattering problem using the Green's tensor formalism

In this appendix, we obtain a solution for the total electromagnetic field, by a separation of the wave equations in a homogeneous part describing the initial electromagnetic field distribution and an inhomogeneous part with a virtual source describing the final electromagnetic field distribution. The Green's tensors required to obtain a solution for these newly formed inhomogeneous wave equations are discussed next.

Appendix B.1. Solution of the scattering problem

To find a solution for the electromagnetic field scattered by structures of a different material with a more complicated shape other than planar interfaces, we separate a virtual source term from an analytically known stratified media term. Starting with Maxwell's equations

$$\nabla \times \mathbf{E} = i\omega\mu\mathbf{H}, \quad (\text{B.1a})$$

$$\nabla \cdot \epsilon\mathbf{E} = \rho_e, \quad (\text{B.1b})$$

$$\nabla \times \mathbf{H} = -i\omega\epsilon\mathbf{E} + \mathbf{J}_e, \quad (\text{B.1c})$$

$$\nabla \cdot \mu\mathbf{H} = 0, \quad (\text{B.1d})$$

we identify as the source term the current density \mathbf{J}_e and charge density ρ_e . Furthermore, we introduce a constant $\kappa = (\mu\epsilon)^{1/2}\omega$ for the wavenumber. Since we are operating in the optical regime, we consider $\mu = \mu_0$. The Maxwell equations combine to

$$\nabla \times \nabla \times \mathbf{E} - \kappa^2 \mathbf{E} = i\omega\mu_0 \mathbf{J}_e, \quad (\text{B.2a})$$

$$\epsilon \nabla \times \frac{1}{\epsilon} \nabla \times \mathbf{H} - \kappa^2 \mathbf{H} = \epsilon \nabla \times \frac{1}{\epsilon} \mathbf{J}_e. \quad (\text{B.2b})$$

Now we identify two situations, described by the set of sourceless Maxwell equations ($\rho_e = 0$, $\mathbf{J}_e = 0$), where a subscript u or t denotes the initial or final situation. It is straightforward to

identify a virtual source term to describe the scattered fields $\mathbf{E}_s = \mathbf{E}_t - \mathbf{E}_u$ and $\mathbf{H}_s = \mathbf{H}_t - \mathbf{H}_u$ in the initial configuration. The source term is obtained as follows:

$$\begin{aligned} \mathbf{J}_e &= i\omega\epsilon_u\mathbf{E}_s + \nabla \times \mathbf{H}_s \\ &= i\omega\epsilon_u(\mathbf{E}_t - \mathbf{E}_u) + \nabla \times (\mathbf{H}_t - \mathbf{H}_u) \\ &= i\omega\epsilon_u(\mathbf{E}_t - \mathbf{E}_u) - i\omega(\epsilon_t\mathbf{E}_t - \epsilon_u\mathbf{E}_u) \\ &= -i\omega(\epsilon_t - \epsilon_u)\mathbf{E}_t. \end{aligned} \quad (\text{B.3})$$

The wave equations describing the system given in equations (B.2) with the virtual source extension combine to

$$\nabla \times \nabla \times \mathbf{E}_s - \kappa_u^2\mathbf{E}_s = (\kappa_t^2 - \kappa_u^2)\mathbf{E}_t, \quad (\text{B.4a})$$

$$\epsilon_u \nabla \times \frac{1}{\epsilon_u} \nabla \times \mathbf{H}_s - \kappa_u^2\mathbf{H}_s = -i\omega\epsilon_u \nabla \times \left(\frac{\epsilon_t - \epsilon_u}{\epsilon_u} \right) \mathbf{E}_t. \quad (\text{B.4b})$$

Since equations (B.4a) and (B.4b) depend on the final electric field, we have to solve an inverse problem. To obtain a solution, we need to know either the Green's tensor, describing the dipole response or the electromagnetic field in the final situation. We use the latter condition, in combination with the set of inhomogeneous wave equations for the general electric source, equation (B.3), and the electromagnetic field and permittivity distribution of the initial configuration denoted by subscript u . Next, we apply a corollary of Green's second theorem, [70], modified for tensors, to derive an expression for the electric field,

$$\mathbf{E}_s(\mathbf{r}') = \int (\epsilon_t - \epsilon_u) \mathbf{G}_e^T(\mathbf{r}, \mathbf{r}') \cdot \mathbf{E}_t(\mathbf{r}) d\mathbf{r}, \quad (\text{B.5a})$$

$$\mathbf{H}_s(\mathbf{r}') = \int (\epsilon_t - \epsilon_u) \mathbf{G}_h^T(\mathbf{r}, \mathbf{r}') \cdot \mathbf{E}_t(\mathbf{r}) d\mathbf{r}, \quad (\text{B.5b})$$

where the integration takes part over the total volume \mathbb{R}^3 but is only non-zero inside the scattering structures. An alternative solution is obtained in the form of surface integrals:

$$\mathbf{E}_s(\mathbf{r}') = \frac{i}{\omega} \oint [\mathbf{G}_h^T(\mathbf{r}, \mathbf{r}') \times \mathbf{E}_s(\mathbf{r}) + \mathbf{G}_e^T(\mathbf{r}, \mathbf{r}') \times \mathbf{H}_s(\mathbf{r})] \cdot \hat{\mathbf{n}} d\mathbf{A}_r, \quad (\text{B.6a})$$

$$\mathbf{H}_s(\mathbf{r}') = \frac{i}{\omega} \oint \left[\mathbf{G}_h^T(\mathbf{r}, \mathbf{r}') \times \mathbf{H}_s(\mathbf{r}) - \frac{\epsilon_u}{\mu_0} \mathbf{G}_e^T(\mathbf{r}, \mathbf{r}') \times \mathbf{E}_s(\mathbf{r}) \right] \cdot \hat{\mathbf{n}} d\mathbf{A}_r. \quad (\text{B.6b})$$

where the integration surface encloses all sources and $\hat{\mathbf{n}}$ is the surface normal pointing inwards. Finally, the inverse problem in equations (B.5a) and (B.5b) or (B.6a) and (B.6b) has to be solved. First the scatterer will be discretized by replacing the integral with summation over a finite amount of cells. Next, the inverse problem is solved using an iterative procedure as described in [69].

Appendix B.2. The Green's tensor formalism

In this section we derive a solution for the Maxwell equations, describing the electromagnetic response from an electric dipole source, as schematically depicted in figure B1:

$$\nabla \times \mathbf{G}_e(\mathbf{r}, \mathbf{r}') = i\omega\mu\mathbf{G}_h(\mathbf{r}, \mathbf{r}'), \quad (\text{B.7a})$$

$$\nabla \cdot \epsilon\mathbf{G}_e(\mathbf{r}, \mathbf{r}') = -\nabla \cdot \delta(\mathbf{r} - \mathbf{r}')\mathbf{I}, \quad (\text{B.7b})$$

$$\nabla \times \mathbf{G}_h(\mathbf{r}, \mathbf{r}') = -i\omega\epsilon\mathbf{G}_e(\mathbf{r}, \mathbf{r}') - i\omega\delta(\mathbf{r} - \mathbf{r}')\mathbf{I}, \quad (\text{B.7c})$$

$$\nabla \cdot \mu\mathbf{G}_h(\mathbf{r}, \mathbf{r}') = 0. \quad (\text{B.7d})$$

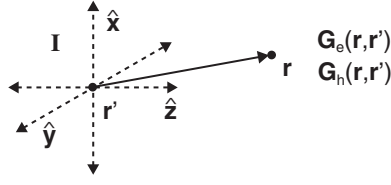


Figure B1. Green's tensors at \mathbf{r} obtained for an electric source \mathbf{G}_e and \mathbf{G}_h located at \mathbf{r}' . The tensor consists of the electric or magnetic field response from a source oscillating in the x -, y - and z -directions.

Once more we restrict ourselves to $\mu = \mu_0$ and find for the wave equations

$$\nabla \times \nabla \times \mathbf{G}_e - \kappa^2 \mathbf{G}_e = \omega^2 \mu_0 \delta(\mathbf{r} - \mathbf{r}') \mathbf{I}, \quad (\text{B.8a})$$

$$\epsilon \nabla \times \frac{1}{\epsilon} \nabla \times \mathbf{G}_h - \kappa^2 \mathbf{G}_h = -i\omega \epsilon \nabla \times \frac{\delta(\mathbf{r} - \mathbf{r}')}{\epsilon} \mathbf{I}. \quad (\text{B.8b})$$

Next, we apply the tensor identity, $\nabla \times \nabla \times \mathbf{A} = \nabla \nabla \cdot \mathbf{A} - \nabla^2 \mathbf{A}$, to obtain

$$\nabla^2 \mathbf{G}_e + \kappa^2 \mathbf{G}_e = -\omega^2 \mu_0 \delta(\mathbf{r} - \mathbf{r}') \mathbf{I} + \nabla (\nabla \cdot \mathbf{G}_e), \quad (\text{B.9a})$$

$$\nabla^2 \mathbf{G}_h + \kappa^2 \mathbf{G}_h = i\omega \nabla \times \delta(\mathbf{r} - \mathbf{r}') \mathbf{I} + \nabla (\nabla \cdot \mathbf{G}_h), \quad (\text{B.9b})$$

By substituting the divergence of the Green's tensors, as given in equations (B.7a)–(B.7d), we obtain

$$\nabla^2 \mathbf{G}_e + \kappa^2 \mathbf{G}_e = -\omega^2 \mu_0 \left(\mathbf{I} + \frac{\nabla \nabla}{\kappa^2} \right) \cdot \delta(\mathbf{r} - \mathbf{r}') \mathbf{I}, \quad (\text{B.10a})$$

$$\nabla^2 \mathbf{G}_h + \kappa^2 \mathbf{G}_h = i\omega \nabla \times \delta(\mathbf{r} - \mathbf{r}') \mathbf{I}. \quad (\text{B.10b})$$

To solve this set of inhomogeneous wave equations, we use the solution of the scalar wave equation for a dipole source at location \mathbf{r}' , denoted as the scalar Green's function $g(\mathbf{r}, \mathbf{r}')$,

$$\nabla^2 g(\mathbf{r}, \mathbf{r}') + \kappa^2 g(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'). \quad (\text{B.11})$$

For the one-dimensional (1D), the two-dimensional (2D) and the three-dimensional (3D) situations, respectively, the Green's function is given by

$$g_{1d}(\mathbf{r}, \mathbf{r}') = \frac{i}{2k_z} e^{ik_z|z-z'|} e^{i(k_x x + k_y y)}, \quad (\text{B.12a})$$

$$g_{2d}(\mathbf{r}, \mathbf{r}') = \frac{i}{4} H_0(\rho k_r) e^{ik_y y}, \quad (\text{B.12b})$$

$$g_{3d}(\mathbf{r}, \mathbf{r}') = \frac{e^{i\kappa|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}. \quad (\text{B.12c})$$

For the 1D-case the source extends to infinity in the x and y directions, and both the x - and the y -component of the wave vector, k_x and k_y , are constant and satisfy the relation $\kappa^2 = k_x^2 + k_y^2 + k_z^2$. For the 2D-case the source extends to infinity in the y direction, the effective distance to the line-source is given by the cylindrical coordinate ρ and the y -component of the wave vector, k_y , is constant and satisfies the relation $\kappa^2 = k_r^2 + k_y^2$. The function $H_n(\rho k_r) = J_n(\rho k_r) + i Y_n(\rho k_r)$ denotes the Hankel function of the first kind, of order n .

To solve the set of equations (B.10), we use once more the corollary of Green's second theorem, [70]. Note that the integration should be carried out carefully if the integration volume includes the point $\mathbf{r} = \mathbf{r}'$, since Green's function is singular in that point. The explicit solution

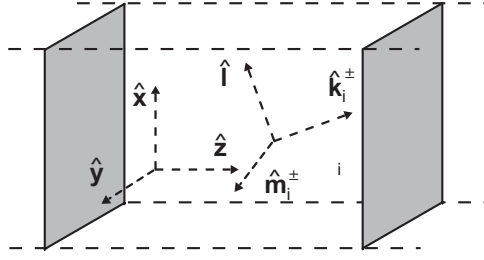


Figure B2. System-specific choice of a medium-dependent set of vectors forming an orthonormal basis, a material transition in z -direction is included in the system. The set of vectors $(\hat{\mathbf{k}}_i^\pm, \hat{\mathbf{l}}_i^\pm, \hat{\mathbf{m}}_i^\pm)$ corresponds to the direction of propagation, the s -polarization and the p -polarization, respectively, the superscript \pm denotes forward or backward propagation in the z -direction.

for the Green's tensors at $\mathbf{r} \neq \mathbf{r}'$ in a single medium is given by

$$\begin{aligned} \epsilon \mathbf{G}_e(\mathbf{r}, \mathbf{r}') &= \kappa^2 \left(\mathbf{I} + \frac{1}{\kappa^2} \nabla \nabla \right) g(\mathbf{r}, \mathbf{r}') \\ &= \kappa^2 \left[\frac{i\kappa R - 1}{\kappa^2 R^2} (\mathbf{I} - 3\hat{\mathbf{R}}\hat{\mathbf{R}}) + (\mathbf{I} - \hat{\mathbf{R}}\hat{\mathbf{R}}) \right] g(\mathbf{r}, \mathbf{r}'), \end{aligned} \quad (\text{B.13a})$$

$$\begin{aligned} \mathbf{G}_h(\mathbf{r}, \mathbf{r}') &= -i\omega \nabla \times g(\mathbf{r}, \mathbf{r}') \mathbf{I} \\ &= -i\omega \frac{i\kappa R - 1}{R} (\hat{\mathbf{R}} \times \mathbf{I}) g(\mathbf{r}, \mathbf{r}'), \end{aligned} \quad (\text{B.13b})$$

where we defined $\mathbf{R} = \mathbf{r} - \mathbf{r}'$, and $\hat{\mathbf{R}}\hat{\mathbf{R}}$ denotes the matrix with elements $R_i R_j$.

To obtain the solution for the Green's tensors in stratified media, from equations (B.8a) and (B.8b), it will prove useful to take the Fourier transform, yielding

$$\epsilon \mathcal{F}[\mathbf{G}_e(\mathbf{k}, \kappa)] = \frac{\kappa^2 \mathbf{I} - \mathbf{k}\mathbf{k}}{k^2 - \kappa^2}, \quad (\text{B.14a})$$

$$\mathcal{F}[\mathbf{G}_h(\mathbf{k}, \kappa)] = -i\omega \frac{i\mathbf{k} \times \mathbf{I}}{k^2 - \kappa^2}. \quad (\text{B.14b})$$

The coordinates \mathbf{k} in the frequency domain form a Fourier pair with the coordinates \mathbf{r} in the spatial domain. An expression for the Green's tensors in the spatial domain is obtained by taking the inverse Fourier transform of equations (B.14a) and (B.14b):

$$\epsilon \mathbf{G}_e(\mathbf{r}, \mathbf{r}') = \frac{1}{(2\pi)^3} \int \frac{\kappa^2 \mathbf{I} - \mathbf{k}\mathbf{k}}{k^2 - \kappa^2} e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} d\mathbf{k}, \quad (\text{B.15a})$$

$$\mathbf{G}_h(\mathbf{r}, \mathbf{r}') = \frac{\omega}{(2\pi)^3} \int \frac{\mathbf{k} \times \mathbf{I}}{k^2 - \kappa^2} e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} d\mathbf{k}. \quad (\text{B.15b})$$

Appendix B.3. Stratified media

So far we concentrated on the calculation of the Green's tensor in a single homogeneous medium. In this section we give a description of the Green's tensors for the general situation of stratified media following [69]. In practice we separate the Green's tensor in a direct and an indirect contribution, which is the light travelling from the source point to the observation point directly, without encountering any material transitions, and the part affected by medium transitions and reflections.

Again, we assume a stack of layers with the surface normal pointing in the z -direction. To describe the effects of the layer transitions on the Green's tensors, we use a similar approach

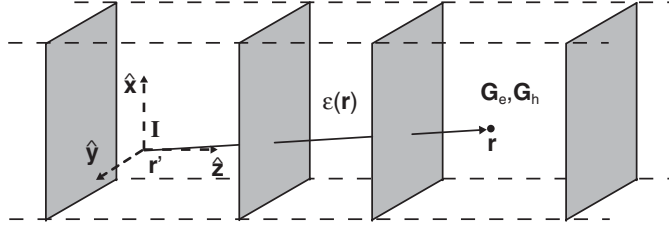


Figure B3. Green's tensors for stratified media as obtained by an electric G_e and G_h . The tensors are constructed from the three vectors containing the electric or magnetic field as obtained by a source oscillating in the x -, y - and z -direction. The source is located at \mathbf{r}' , the point of observation is located at \mathbf{r} .

as discussed in [appendix A.2](#) and introduce the cylindrical basis $(\hat{k}, \hat{l}, \hat{m})$ with the three unit vectors defined in equation (A.5), as shown in figure B2. We rewrite equations (B.15a) and (B.15b) in terms of our new basis-vectors, starting in the layer i where the source \mathbf{r}' is located:

$$\begin{aligned}
 \epsilon_i \mathbf{G}_e(\mathbf{r}, \mathbf{r}') &= \frac{1}{(2\pi)^3} \int \frac{\kappa_i^2 \mathbf{I} - \mathbf{k}\mathbf{k}}{k^2 - \kappa_i^2} e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} d\mathbf{k} \\
 &= \frac{1}{(2\pi)^3} \int \left(\hat{z}\hat{z} + \frac{\kappa_i^2 \mathbf{I} - \mathbf{k}\mathbf{k}}{k^2 - \kappa_i^2} \right) e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} d\mathbf{k} - \frac{1}{(2\pi)^3} \int \hat{z}\hat{z} e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{k}')} d\mathbf{k} \\
 &= \frac{i}{8\pi^2} \int \frac{\kappa_i^2 \mathbf{I} - \mathbf{k}_i^\pm \mathbf{k}_i^\pm}{k_{zi}} e^{i\mathbf{k}_i^\pm \cdot (\mathbf{r} - \mathbf{r}')} k_r dk_r dk_\phi - \delta(\mathbf{r} - \mathbf{r}') \hat{z}\hat{z} \\
 &= -\delta(\mathbf{r} - \mathbf{r}') \hat{z}\hat{z} + \frac{i\kappa_i^2}{8\pi^2} \int \frac{\hat{\mathbf{u}} + \hat{\mathbf{m}}_i^\pm \hat{\mathbf{m}}_i^\pm}{k_{zi}} e^{i\mathbf{k}_i^\pm \cdot (\mathbf{r} - \mathbf{r}')} k_r dk_r dk_\phi,
 \end{aligned} \tag{B.16a}$$

$$\begin{aligned}
 \mathbf{G}_h(\mathbf{r}, \mathbf{r}') &= \frac{\omega}{(2\pi)^3} \int \frac{\mathbf{k} \times \mathbf{I}}{k^2 - \kappa_i^2} e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} d\mathbf{k} \\
 &= \frac{\omega}{(2\pi)^3} \int \frac{\mathbf{k} \times (\hat{k}\hat{k} + \hat{\mathbf{u}} + \hat{\mathbf{m}}\hat{\mathbf{m}})}{k^2 - \kappa_i^2} e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} d\mathbf{k} \\
 &= \frac{i\omega}{8\pi^2} \int \frac{\mathbf{k}_i^\pm \times (\hat{\mathbf{u}} + \hat{\mathbf{m}}_i^\pm \hat{\mathbf{m}}_i^\pm)}{k_{zi}} e^{i\mathbf{k}_i^\pm \cdot (\mathbf{r} - \mathbf{r}')} k_r dk_r dk_\phi \\
 &= \frac{i\omega}{8\pi^2} \int \kappa_i \frac{\hat{\mathbf{m}}_i^\pm \hat{\mathbf{l}} - \hat{\mathbf{l}} \hat{\mathbf{m}}_i^\pm}{k_{zi}} e^{i\mathbf{k}_i^\pm \cdot (\mathbf{r} - \mathbf{r}')} k_r dk_r dk_\phi,
 \end{aligned} \tag{B.16b}$$

where, after integration over k_z , the length of the vector $|\mathbf{k}_i^\pm| = \kappa_i$, the positive sign is used for $z > z'$ and the negative sign is used for $z < z'$. For the integration over k_z in equations (B.16a) and (B.16b), we have anticipated a singularity caused by the layer transitions and taken this component out of the equation, where it is integrated separately.

Now, we incorporate the effects of a system consisting of several layers, as schematically depicted in figure B3. This yields a direct contribution when source and observation point are located within the same medium; otherwise there is only an indirect contribution from possibly multiple reflections and transmissions of the surrounding layers. The electric and magnetic field components incorporating the reflection and transmission coefficients of the

layer transitions are given by

$$\begin{aligned}
\mathbf{G}_e(\mathbf{r}, \mathbf{r}') &= -\frac{\delta(\mathbf{r} - \mathbf{r}')}{\epsilon_i} \hat{\mathbf{z}}\hat{\mathbf{z}} + \frac{i\omega^2\mu_0}{8\pi^2} \int \frac{1}{k_{zi}} \left[\hat{\mathbf{u}} e^{ik_{zi}|z-z'|} + \mathbf{R}_i^{s+} e^{ik_{zi}z} + \mathbf{R}_i^{s-} e^{-ik_{zi}z} \right. \\
&\quad \left. + \hat{\mathbf{m}}_i^\pm \hat{\mathbf{m}}_i^\pm e^{ik_{zi}|z-z'|} + \mathbf{R}_i^{p+} e^{ik_{zi}z} + \mathbf{R}_i^{p-} e^{-ik_{zi}z} \right]^{i \rightarrow j} e^{i\rho k_r \cos(k_\phi - \varphi)} k_r dk_r dk_\phi \\
&= -\frac{\delta(\mathbf{r} - \mathbf{r}')}{\epsilon_i} \hat{\mathbf{z}}\hat{\mathbf{z}} + \frac{i\omega^2\mu_0}{8\pi^2} \int \left[(\mathbf{A}_j^{s+} + \mathbf{A}_j^{p+}) e^{ik_{zj}z} \right. \\
&\quad \left. + (\mathbf{A}_j^{s-} + \mathbf{A}_j^{p-}) e^{-ik_{zj}z} \right] e^{i\rho k_r \cos(k_\phi - \varphi)} \frac{k_r}{k_{zi}} dk_r dk_\phi, \tag{B.17a}
\end{aligned}$$

$$\begin{aligned}
\mathbf{G}_h(\mathbf{r}, \mathbf{r}') &= \frac{i\omega}{8\pi^2} \int \mathbf{k}_j^\pm \times \frac{1}{k_{zi}} \left[\hat{\mathbf{u}} e^{ik_{zi}|z-z'|} + \mathbf{R}_i^{s+} e^{ik_{zi}z} + \mathbf{R}_i^{s-} e^{-ik_{zi}z} \right. \\
&\quad \left. + \hat{\mathbf{m}}_i^\pm \hat{\mathbf{m}}_i^\pm e^{ik_{zi}|z-z'|} + \mathbf{R}_i^{p+} e^{ik_{zi}z} + \mathbf{R}_i^{p-} e^{-ik_{zi}z} \right]^{i \rightarrow j} e^{i\rho k_r \cos(k_\phi - \varphi)} k_r dk_r dk_\phi \\
&= \frac{i\omega}{8\pi^2} \int \left[\mathbf{k}_j^+ \times (\mathbf{A}_j^{s+} + \mathbf{A}_j^{p+}) e^{ik_{zj}z} \right. \\
&\quad \left. + \mathbf{k}_j^- \times (\mathbf{A}_j^{s-} + \mathbf{A}_j^{p-}) e^{-ik_{zj}z} \right] e^{i\rho k_r \cos(k_\phi - \varphi)} \frac{k_r}{k_{zi}} dk_r dk_\phi, \tag{B.17b}
\end{aligned}$$

with the vector $\mathbf{r} - \mathbf{r}'$ expressed in cylindrical coordinates as (ρ, φ, z) and $\mathbf{R}_i^{s/p+}$ and $\mathbf{R}_i^{s/p-}$ the effective reflection tensors for electric waves travelling in the positive and negative direction, respectively. We have introduced the layer-dependent electric field components $\mathbf{A}_j^{s/p\pm}$, where the positive sign denotes forward and the negative sign backward propagating fields. The arrow $i \rightarrow j$ indicates that we start in the layer where the source point is located and migrate to the layer where the observation point is located. The combination of direct and indirect waves are stored in $\mathbf{A}_i^{s/p\pm}$ as

$$\mathbf{A}_i^{s/p\pm} = \left[\mathbf{R}_i^{s/p\pm} + \Theta(\pm z \mp z') \mathbf{C}_i^{s/p\pm} e^{\mp ik_{zi}z'} \right], \tag{B.18}$$

with $\mathbf{C}_i^{s\pm} = \hat{\mathbf{u}}$ for the TE- and $\mathbf{C}_i^{p\pm} = \hat{\mathbf{m}}_i^\pm \hat{\mathbf{m}}_i^\pm$ for the TM-polarization. The Heaviside step function $\Theta(x) = 1$ for $x > 0$ and $\Theta(x) = 0$ for $x < 0$.

We can obtain an expression for $\mathbf{R}_i^{s/p\pm}$, if we use for each component of the tensors the relation

$$\left(\frac{\mathbf{A}_i^\pm}{\mathbf{A}_i^\mp} \right)_{z \lesseqgtr z'}^{s/p} = \frac{\mathbf{R}_i^{s/p\pm}}{\mathbf{R}_i^{s/p\mp} + \mathbf{C}_i^{s/p\mp} e^{\pm ik_{zi}z'}}, \tag{B.19}$$

yielding a set of two equations for $z \lesseqgtr z'$, where $z < z'$ is associated with the ratio of positive over negative waves and $z > z'$ is associated with the ratio of negative over positive waves. For each element of the effective reflection tensors we find

$$\mathbf{R}_{i,\alpha\beta}^{s/p\pm} = \frac{(\mathbf{A}_{\alpha\beta}^\pm / \mathbf{A}_{\alpha\beta}^\mp)_{i,z \lesseqgtr z'}^{s/p}}{1 - (\mathbf{A}_{\alpha\beta}^+ / \mathbf{A}_{\alpha\beta}^-)_{i,z \lesseqgtr z'}^{s/p} (\mathbf{A}_{\alpha\beta}^- / \mathbf{A}_{\alpha\beta}^+)_{i,z \gtrless z'}^{s/p}} \left[(\mathbf{A}_{\alpha\beta}^\mp / \mathbf{A}_{\alpha\beta}^\pm)_{i,z \gtrless z'}^{s/p} \mathbf{C}_{i,\alpha\beta}^{s/p\pm} e^{\mp ik_{zi}z'} + \mathbf{C}_{i,\alpha\beta}^{s/p\mp} e^{\pm ik_{zi}z'} \right]. \tag{B.20}$$

Since we know that in the last medium with $i = N$ there are no backward travelling waves coming from infinity, the ratio $(\mathbf{A}^- / \mathbf{A}^+)_{i=N}^{s/p} = 0$ as defined in equation (A.12). In the first medium with $i = 1$, the same holds for the forward over backward travelling waves $(\mathbf{A}^+ / \mathbf{A}^-)_1^{s/p} = 0$. Now, it is possible to calculate this ratio in the adjacent layers until we obtain the ratio in the layer of the source. Note the difference between this ratio and that with subscripts $\alpha\beta$, since the definition of the reflection coefficient for the p -polarization includes a rotation of the vector, which implies an extra minus sign for $\alpha = z$. The calculation of $\mathbf{A}_i^{s/p\pm}$

leads to the expressions

$$\mathbf{A}_i^{s\pm} = \frac{(A^\pm/A^\mp)_i^s}{1 - (A^\pm/A^\mp)_i^s (A^\mp/A^\pm)_i^s} [(A^\mp/A^\pm)_i^s \hat{\mathbf{u}} e^{\mp ik_{zi} z'} + \hat{\mathbf{u}} e^{\pm ik_{zi} z'}] + \hat{\mathbf{u}} \Theta(\pm z \mp z') e^{\mp ik_{zi} z'}, \quad (\text{B.21a})$$

$$\mathbf{A}_i^{p\pm} = \frac{(A^\pm/A^\mp)_i^p}{1 - (A^\pm/A^\mp)_i^p (A^\mp/A^\pm)_i^p} [(A^\mp/A^\pm)_i^p \hat{\mathbf{m}}_i^\pm \hat{\mathbf{m}}_i^\pm e^{\mp ik_{zi} z'} + \hat{\mathbf{m}}_i^\mp \hat{\mathbf{m}}_i^\pm e^{\pm ik_{zi} z'}] + \hat{\mathbf{m}}_i^\pm \hat{\mathbf{m}}_i^\pm \Theta(\pm z \mp z') e^{\mp ik_{zi} z'}, \quad (\text{B.21b})$$

To propagate the field calculated in layer i to the next layer j we use equation (A.14):

$$\mathbf{A}_j^{s/p\pm} = \gamma_{i,j}^{s/p\pm} [A_i^{s/p\pm} e^{\pm ik_{zi} d_{i,j}} - A_i^{s/p\mp} F_{i,j}^{s/p} e^{\mp ik_{zi} d_{i,j}}] e^{\mp ik_{zj} d_{i,j}}. \quad (\text{B.22})$$

Finally, we require, aside from equations (A.18) and (A.19), the expressions for $\mathbf{k}_j^\pm \times \hat{\mathbf{u}} = k_j \hat{\mathbf{m}}_j^\pm \hat{\mathbf{l}}$ and $\mathbf{k}_j^\pm \times \hat{\mathbf{m}}_j^\pm \hat{\mathbf{m}}_i^\pm = -k_j \hat{\mathbf{l}} \hat{\mathbf{m}}_i^\pm$

$$\int_0^{2\pi} \hat{\mathbf{m}}_j^\pm \hat{\mathbf{l}} e^{i\rho k_r \cos(k_\phi - \varphi)} dk_\phi = 2\pi \begin{pmatrix} \mp \frac{k_{zj}}{2} J_2 \sin 2\varphi & \mp \frac{k_{zj}}{2} (J_0 - J_2 \cos 2\varphi) & 0 \\ \pm \frac{k_{zj}}{2} (J_0 + J_2 \cos 2\varphi) & \pm \frac{k_{zj}}{2} J_2 \sin 2\varphi & 0 \\ -ik_r J_1 \sin \varphi & ik_r J_1 \cos \varphi & 0 \end{pmatrix}, \quad (\text{B.23})$$

$$\begin{aligned} & - \int_0^{2\pi} k_j \hat{\mathbf{l}} \hat{\mathbf{m}}_i^\pm e^{i\rho k_r \cos(k_\phi - \varphi)} dk_\phi \\ & = \frac{2\pi k_j}{k_i} \begin{pmatrix} \frac{k_{zi}}{2} J_2 \sin 2\varphi & -\frac{k_{zi}}{2} (J_0 + J_2 \cos 2\varphi) & ik_r J_1 \sin \varphi \\ \frac{k_{zi}}{2} (J_0 - J_2 \cos 2\varphi) & -\frac{k_{zi}}{2} J_2 \sin 2\varphi & -ik_r J_1 \cos \varphi \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned} \quad (\text{B.24})$$

References

- [1] Baird J L 1928 *Improvements in or relating to the recording of views of objects or scenes, or optical images or the like* British Patent GB289104
- [2] Rice P, Macovski A, Jones E D, Frohbach H, Crews R W and Noon A W 1970 An experimental television recording and playback system using photographic discs *J. Soc. Motion Pict. Telev. Eng.* **79** 997
- [3] Welford W T 1972 Relationship between modes of image formation in scanning microscopy and conventional microscopy *J. Microsc.—Oxford* **96** 105
- [4] Born M and Wolf E 1970 *Principles of Optics* (Oxford: Pergamon)
- [5] Sheppard C J R and Choudhury A 1977 Image formation in scanning microscope *Opt. Acta* **24** 1051–73
- [6] Wilson T and Sheppard C J R 1984 *Theory and Practice of the Scanning Optical Microscope* (Cambridge: Cambridge University Press)
- [7] Zernike F 1934 Beugungstheorie des Schneidverfahrens und seiner verbesserten Form, der Phasenkontrastmethode *Physica* **1** 689–704
- [8] Compaan K and Kramer P 1973 The Philips ‘VLP’ System *Philips Tech. Rev.* **33** 178
- [9] Bouwhuis G and Burgstede P 1973 The optical scanning system of the philips ‘VLP’ record player *Philips Tech. Rev.* **33** 186
- [10] Broussaud G, Spitz E, Tinet C and LeCarvenec F 1974 *IEEE Trans. Broadcast Telev. Receivers* **BTR-20** 332
- [11] Hrbek G W 1974 *J. Soc. Motion Pict. Telev. Eng.* **83** 580
- [12] Bouwhuis G and Braat J J M 1978 Video disk player optics *Appl. Opt.* **17** 1993–2000
- [13] Braat J J M and Bouwhuis G 1978 Position sensing in video disk readout *Appl. Opt.* **17** 2013–21
- [14] Braat J J M and Bouwhuis G 1978 Optical video disks with undulating tracks *Appl. Opt.* **17** 2022–8
- [15] Velzel C H F 1978 Laser beam reading of video records *Appl. Opt.* **17** 2029–36
- [16] Korpel A 1978 Simplified diffraction theory of the video disk *Appl. Opt.* **17** 2037–2042
- [17] Hopkins H H 1979 Diffraction theory of laser read-out systems for optical video disks *J. Opt. Soc. Am.* **69** 4–24
- [18] Bouwhuis G and Braat J 1983 Recording and reading of information on optical disks *Applied Optics and Optical Engineering* vol 9, ed R R Shannon and J C Wyant, (New York: Academic)

- [19] Bouwhuis G, Braat J, Huijser A, Pasman J, van Rosmalen G and Schouhamer Immink K 1984 *Principles of Optical Disc Systems* (Bristol: Hilger)
- [20] Isailovic J 1985 *Videodisc and Optical Memory Systems* (Englewood Cliffs, NJ: Prentice-Hall)
- [21] Marchant A B 1990 *Optical Recording* (Reading, MA: Addison Wesley)
- [22] Sincerbox G and Zavislan J (ed) 1992 *Selected Papers on Optical Storage (SPIE Milestone Series)* vol MS 49 (Bellingham: SPIE Optical Engineering Press)
- [23] Schwartz K 1993 *The Physics of Optical Recording* (Berlin: Springer)
- [24] Mansuripur M 1995 *The Physical Principles of Magneto Optical Data Recording* (Cambridge: Cambridge University Press)
- [25] Williams E W 1996 *Textbooks in Electrical and Electronic Engineering Vol 2, The CD-ROM and Optical Recording Systems* (Oxford: Oxford University Press)
- [26] Stan S 1998 *The CD-ROM Drive, a Brief System Description*, (Dordrecht: Kluwer)
- [27] Carriere J, Narayan R, Yeh W-H, Peng C, Khulbe P, Li L, Anderson R, Choi J and Mansuripur M 2000 Principles of optical disk data storage *Progress in Optics* vol 41, ed E Wolf (Amsterdam: Elsevier)
- [28] Braat J J M, Dirksen P and Janssen A J E M 2003 *Diffraction Read-Out of Optical Discs Optical Imaging and Microscopy* ed P Török and F-J Kao (Berlin: Springer)
- [29] Ignatowsky W 1919 Diffraction by a lens of arbitrary aperture *Trans. Opt. Inst.* **1** 1–36
- [30] Richards B and Wolf E 1959 Electromagnetic diffraction in optical systems II. Structure of the image field in an aplanatic system *Proc. R. Soc. A* **253** 358–79
- [31] Mansuripur M 1986 Distribution of light at and near the focus of high-numerical-aperture objectives *J. Opt. Soc. Am. A* **3** 2086–93
- [32] Mansuripur M 1989 Certain computational aspects of vector diffraction problems *J. Opt. Soc. Am. A* **6** 786–805
- [33] Kant R 1993 An analytical solution of vector diffraction for focusing optical systems *J. Mod. Opt.* **40** 337–48
- [34] Kant R 1993 An analytical solution of vector diffraction for focusing optical systems with Seidel aberrations *J. Mod. Opt.* **40** 2293–311
- [35] Török P, Varga P and Booker G R 1995 Electromagnetic diffraction of light focused through a planar interface between materials of mismatched refractive-indexes-structure of the electromagnetic field—1 *J. Opt. Soc. Am. A* **12** 2136–44
- [36] Török P, Varga P and Nemeth G 1995 Analytical solution of the diffraction integrals and interpretation of wave-front distortion when light is focused through a planar interface between materials of mismatched refractive-indexes *J. Opt. Soc. Am. A* **12** 2660–71
- [37] Sheppard C J R and Török P 1997 Efficient calculation of electromagnetic diffraction in optical systems using a multipole expansion *J. Mod. Opt.* **44** 803–19
- [38] Braat J J M, Dirksen P, Janssen A J E M, and van de Nes A S 2003 Extended Nijboer–Zernike representation of the vector field in the focal region of an aberrated high-aperture optical system *J. Opt. Soc. Am. A* **20** 2281–92
- [39] van de Nes A S, Billy L, Pereira S F, Braat J J M 2004 Calculation of the vectorial field distribution in a stratified focal region of a high numerical aperture imaging system *Opt. Express* **12** 1281–93
- [40] Mansfield S M and Kino G S 1990 Solid immersion microscope *Appl. Phys. Lett.* **57** 2615–16
- [41] Terris B D, Mamin H J, Rugar D and Kino G S 1994 Near-field optical data storage using a solid immersion lens *Appl. Phys. Lett.* **65** 388–90
- [42] Zijp F, Lee J-I, Verschuren C A, van den Eerenbeemd J M A and Bruls D M 2006 Improved near-field recording system for first-surface media with an NA = 1.9 solid immersion lens *Japan. J. Appl. Phys.* **45** 1336–40
- [43] Dil J G and Jacobs B A J 1979 Apparent size of reflecting polygonal obstacles of the order of one wavelength *J. Opt. Soc. Am.* **69** 950–60
- [44] Nevière M, Vincent P and Petit R 1974 Theory of conducting gratings and their applications to optics *Nouvelle Revue d'Optique* **5** 65–77
- [45] Maystre D 1978 New general integral theory for dielectric coated gratings *J. Opt. Soc. Am.* **68** 490–5
- [46] Derrick G H, McPhedran R C, Maystre D and Nevière M 1979 Crossed gratings—theory and its applications *Appl. Phys.* **18** 39–52
- [47] Brok J M and Urbach H P 2002 Simulation of polarization effects in diffraction problems of optical recording *J. Mod. Opt.* **49** 1811–29
- [48] Brok J M and Urbach H P 2003 Rigorous model of the scattering of a focused spot by a grating and its application in optical recording *J. Opt. Soc. Am. A* **20** 256–72
- [49] Hill B and Schmidt K P 1973 New page-composer for holographic data storage *Appl. Opt.* **12** 1193–8
- [50] Nelson R H, Lugt A V and Zech R G 1974 Holographic data storage and retrieval *Opt. Eng.* **13** 429–34
- [51] Burr G W, Jefferson C M, Coufal H, Jurich M, Hoffnagle J A, Macfarlane R M and Shelby R M 2001 Volume holographic data storage at areal density of 250 gigapixels/in.(2) *Opt. Lett.* **26** 444–6

- [52] Yao B L, Ren Z W, Menke N, Wang L Y, Zheng Y, Lei M, Chen G F and Hampp N 2005 Polarization holographic high-density optical data storage in bacteriorhodopsin film *Appl. Opt.* **44** 7344–8
- [53] Bouwhuis G and Spruit J 1990 Optical storage read-out of nonlinear disks *Appl. Opt.* **29** 3766–8
- [54] Tominaga J, Fuji H, Sato A, Nakano T, Fukaya T and Atoda N J 1998 The near-field super-resolution properties of an antimony thin film *Japan. J. Appl. Phys. Part 2—Lett.* **37** L1323–5
- [55] Tominaga J, Fuji H, Sato A, Nakano T and Atoda N J 2000 The characteristics and the potential of super resolution near-field structure *Japan. J. Appl. Phys. Part 1—Regular Papers, Short Notes Review Papers* **39** 957–61
- [56] Awano H, Ohnuki S, Shirai H, Ohta N, Yamaguchi A, Sumi S and Torazawa K 1996 Magnetic domain expansion readout for amplification of an ultra high density magneto-optical recording signal *Appl. Phys. Lett.* **69** 4257–9
- [57] Stamnes J J 1986 *Waves in Focal Regions (Series on Optics and Optoelectronics)* ed E R Pike and W T Welford (Bristol: Hilger)
- [58] van de Nes A S, Pereira S F and Braat J J M 2006 On the conservation of fundamental optical quantities in non-paraxial imaging systems *J. Mod. Opt.* **53** 677–87
- [59] Verschuren C A and van den Eerenbeemd J M A, Zipp F, Lee J-I and Bruls D M 2006 Near-field recording with a solid immersion lens on polymer cover-layer protected discs *Japan. J. Appl. Phys.* **45** 1325–31
- [60] O’Neill M P and Wong T L 2000 Multi-level data storage system using phase-change optical discs *Optical Data Storage (2000 Conference Digest)* pp 170–2
- [61] Dorn R, Quabis S and Leuchs G 2003 Sharper focus for a radially polarized light beam *Phys. Rev. Lett.* **91** 233901
- [62] Stalder M and Schadt M 1996 Linearly polarized light with axial symmetry generated by liquid-crystal polarization converters *Opt. Lett.* **21** 1948–50
- [63] Coene W M J Nonlinear signal-processing model for scalar diffraction in optical recording *Appl. Opt.* **42** 6525–35
- [64] Beijersbergen M W, Coerwinkel R P C, Kristensen M and Woerdman J P 1994 Helical-wavefront laser beams produced with a spiral phaseplate *Opt. Commun.* **112** 321–7
- [65] Siegman A E 1986 *Lasers* (Mill Valley: University Science Books)
- [66] Paulus M, Cay-Balmaz P and Martin O J F 2000 Accurate and efficient computation of the Green’s tensor for stratified media *Phys. Rev. E* **62** 5797–807
- [67] Braat J J M, Dirksen P, Janssen A J E M, van Haver S and van de Nes A S 2005 Extended Nijboer–Zernike approach to aberration and birefringence retrieval in a high-numerical-aperture optical system *J. Opt. Soc. Am. A* **22** 2635–50
- [68] Stallinga S 2004 Strehl ratio for focusing into biaxially birefringent media *J. Opt. Soc. Am. A* **21** 1785–98
- [69] Martin O J F, Girard C and Dereux A 1995 Generalized field propagator for electromagnetic scattering and light confinement *Phys. Rev. Lett.* **74** 526–9
- [70] Jackson J D 1999 *Classical Electrodynamics* (New York: Wiley)