Diffractive read-out of optical discs

Joseph Braat

Optics Research Group
Faculty of Applied Sciences
Delft University of Technology
Lorentzweg 1
NL - 2627 CA Delft
The Netherlands
E-mail: j.j.m.braat@tnw.tudelft.nl

P. Dirksen

Philips Research Laboratories
High Tech Campus 4
Professor Holstlaan 4
NL - 5656 AA Eindhoven
The Netherlands
E-mail: peter.dirksen@philips.com

Augustus J.E.M. Janssen

Philips Research Laboratories
High Tech Campus 36
Professor Holstlaan 4
NL - 5656 AA Eindhoven
The Netherlands
E-mail: a.j.e.m.janssen@philips.com

Summary. We present the optical principles of actual and projected systems for optical data storage like the CD (Compact Disc) , the DVD (Digital Versatile Disc) and the blue light DVR system. After a short historic overview, the initial choices with respect to optical read-out method, light path and optical components are briefly reviewed. We next consider in more detail the radial tracking method applied in the DVD-system and the compatibility aspects between the DVD- and CD-system. Throughout this chapter we especially pay attention to the modelling of the signal read-out in an optical disc player and we will describe methods to efficiently calculate crucial parameters like signal jitter and cross-talk.

1 Introduction

An extensive literature is available on optical data storage. Because of the multi-disciplinary character of optical storage, the subjects range from optics, mechanics, control theory, electronics, signal coding and cryptography to chemistry and solid-state physics. Several books [2]-[9] have appeared that are devoted to one or several of the subjects mentioned above. This chapter will be limited to purely optical aspects of the optical storage systems. In Section 2 of this chapter we first present a short historic overview of the research on optical disc systems that has led to the former video disc system. In Section 3 we briefly review the general principles of optical storage systems that have remained more or less unchanged since the start some thirty years ago. An interesting feature of the DVD-system is its standardised radial tracking method that we will treat in some detail in Section 4. The appearance of new generations with higher spatial density and storage capacity has triggered solutions for the backward compatibility of the new systems with the existing ones. In Section 5 we pay attention to these backward compatibility problems that are mainly caused by the smaller cover layer thickness, the shorter wavelength of the laser source and the higher numerical aperture of the objective in the new generations. Finally, in Section 6 we indicate how an efficient modelling of the optical read-out system can be done. Important features like the radial cross-talk and inter-symbol interference can be studied with an advanced model of the optical read-out and the resulting signal jitter, an important quality factor for a digital signal, is efficiently obtained. A new approach is presented that can speed up the calculations so that the modelling can really serve as an interactive tool when optimizing the optical read-out.

2 Historic overview of video and audio recording on optical media

Optical data storage uses the principle of the point-by-point scanning of an object to extract the recorded information on the disc. In another context, such an approach was used in a somewhat primitive way in the 'flying-spot' film scanner. This apparatus was developed in the early fifties of the last century when a need was felt to transform the pictorial information on film into electrical information to be recorded on magnetic tape. The flying-spot in the film scanner was obtained with the aid of a cathode ray tube and its resolving power was adequate for the resolution encountered on the standard cinema film format. In Fig.(1) the focusing lens L_o is shown that produces the scanning spot on the film. The light is partially transmitted and diffracted by the information on the film surface and the detector collects the transmitted light via the collecting lens L_c. The required resolution in the film plane was of the order of 20 to 30 μm and this was not an extreme demand neither for the electron optics nor for the 'objective' lens L_o. The flying-spot scanner was not analysed in depth for some time, it just functioned correctly for the purpose it had been developed for and produced coloured images with the required TV-resolution.

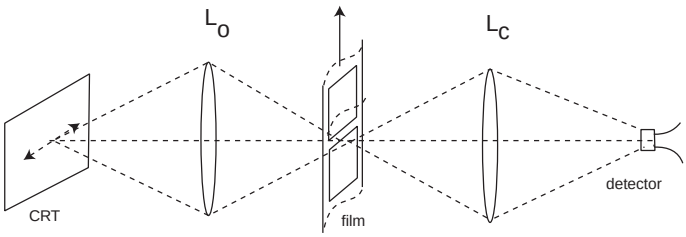

Fig. 1. Schematic lay-out of a flying-spot film scanning device.

In 1961 a patent was attributed to M. Minsky [10] where he describes the principle of a scanning microscope with a resolving power comparable to the standard microscope. Minsky even proposed to use several scanning spots in parallel, thus realising a large-field scanning . The practical use of his (confocal) scanning microscope was limited because of the inefficient use of the power of the (thermal) light source.
A further step towards the optical data storage principle as we know it nowadays was taken by a research group at Stanford University that focused on the recording of video information on a record with the size of a standard audio Long Play disc [11]. The group developed a disc-shaped carrier with data tracks that carried the directly modulated video signal. After a few years the research was abandoned because of the cross-talk problem between the information in neighbouring tracks and because of the bad signal to noise ratio of the detected signal. The belief in a commercial success as a consumer product dwindled away and the sponsor, the 3M-company, stopped funding.
In parallel to this early research on direct signal recording, several groups worked on the principle of storing pictures on a disc, either directly or via holographic encoding. Of course, the latter method was believed to be much more robust because of the far-field recording of the pictorial information. However, one has to conclude that the various prototypes of 'pictorial' video-players that were developed at the end of the sixties and the early 1970's were not satisfactory because of the inefficient use of the available recording surface, the difficult mechanics in the case of the 'direct picture' approach and the colour and speckle problem for the holographic devices.

2.1 The early optical video system

At the end of the 1960's, various disciplines and techniques were present at Philips Research Laboratories in the Netherlands that enabled a breakthrough in the field of optical data storage. Important factors enabling this breakthrough were

- availability of expertise in precision optics and control mechanics,
- research on a consumer-type of HeNe-laser where the delicate problems of alignment and stability of such a coherent source were solved in such a way that relatively cheap mass-production could be envisaged,
- work on efficient modulation techniques for video recording on magnetic tape,

- a strong belief in the market for educational applications (teaching, instruction, continuous learning).

In an early stage, some decisions [12] were taken (reflective read-out, information protection via the disc substrate) that have proven to be of vital importance for the success of 'robust' optical data storage. In 1975, several companies (Philips, Thomson [13], Music Corporation of America [14]), later joined by Pioneer, have united to establish a standard for the video disc system (VLP = Video Long Play) so that worldwide disc exchange would be possible. The most important standardised features were

- HeNe laser source, λ=633 nm, numerical aperture (NA) of the read-out objective equal to 0.40, track pitch of information spiral is 1.6 µm, disc diameter is 30 cm, rotational velocity is 25 or 30 Hz depending on the television standard.
- frequency modulation of the analog video signal resulting in optical information pits with a (small) pit length modulation; the average pit length typically is 0.6 µm with comparable 'land' regions in between.
- recording of two video frames per revolution (constant angular velocity, CAV) or recording with constant linear velocity (CLV). The first choice allowed for the still-picture option without using a frame memory (not yet available at that time), the second choice yielded discs with the longest playing time (up to two hours)

Further research on the video disc system (later called the Laser Disc system) has centred on an increase of the recording density, e.g. by

- halving the track pitch to 0.8 µm and keeping the cross-talk at an acceptable level by introducing a two-level pit depth that alternates from track to track [15]
- using (high-frequency) track undulation for storing extra information, e.g. colour and/or sound channels

Other research efforts have resulted in the introduction of the semiconductor laser (λ=820 nm) and the simplification of the optics, among others by introducing aspherical optics for the scanning objective [16].
Further potential changes in disc definition resulting from this research on higher density have not led to an improved video disc standard; but the semiconductor laser source and the cheaper optics have been readily introduced at the beginning of the 1980's.
In parallel to the video application, storage of purely digital data has been pursued in the early stages of optical storage. Various systems using large-size optical discs have been put on the professional market while the recording of data was made possible by the development of high-power semiconductor lasers. In the first instance, the data were irreversibly 'burnt' in the recording layer. Later research has revealed the existence of suitable phase-change materials [17],[18] that show an optical contrast between the amorphous and crystalline phase and permit rapid thermally induced switching between both phases. Other optical recording materials rely on the laser-induced thermal switching of magnetic domains [7]; the optical read-out of

the domains with opposite vertical magnetisation is done using the magneto-optical Kerr effect.

2.2 The origin of the CD-system

The Laser Disc video system knew a limited success due to its restricted playing time and its lack of the recording option which made direct competition with the well-established video cassette recorder rather difficult. Other opportunities for the optical storage technique were already examined for several years and the replacement of the audio Long Play disc by an optical medium was actively researched. An enormous extra capacity becomes available when switching from the mechanical LP needle to the micron-sized diffraction-limited optical 'stylus'. The approximately hundred times larger storage capacity on an equally sized optical disc has been used in two ways, first by reducing the size of the new medium and, secondly, by turning towards digital encoding of the audio signal which, at that time, asked for a very substantial increase in signal bandwidth. After a certain number of iterations where disc size, signal modulation method, digital error correction scheme and playing time were optimised, the actual Compact Disc standard was established in 1980 by the partners Philips and Sony. The disc diameter of 12 cm was established to accommodate for the integral recording on a single disc of the longest version of the ninth symphony of Beethoven (74 minutes playing time, Bayreuth 1951 recording by the Berliner Philharmoniker, conductor Wilhelm Fuertwangler).
The worldwide market introduction of the CD-system took place at the end of 1982 and the beginning of 1983. Apart from the infrared semiconductor laser source and the corresponding miniaturisation of the optics, no really fundamental changes with respect to the preceding laser disc video system were present regarding the optical read-out principles. The CD system turned out to be simpler and more robust than its video signal predecessor. This was mainly due to the digitally encoded signal on the CD-disc that is much less sensitive to inter-track cross-talk than the former analog video signal.

2.3 The road towards the DVD-system

For a long period, the application of the CD-system has been restricted to the digital audio domain. In an early stage, a read-only digital data storage system with identical capacity (650 MByte) has been defined (CD-ROM) but the market penetration of this system has been relatively late, some 8 to 10 years after the introduction of the audio CD. Once the personal computer application took off, the demand for recordable and rewritable CD-systems (CD-R and CD-RW) immediately followed. At the same time, the computer environment asked for an ever increasing data capacity to handle combination of pictorial and data content (e.g. for games). Spurred by the quickly developing market of video (using the lower quality digital MPEG1 video standard), a serious need was felt for an improved optical data storage standard. The DVD-standard has been defined as of 1994 by a consortium of 10 companies and offers a seven times higher capacity than the CD-format. Because of its multi-purpose

application, the acronym DVD stands for Digital Versatile Disc and both MPEG2 video information, super-audio signals and purely digital data can be recorded on this medium. Again, the optical principles used in the DVD-system have not fundamentally changed; a relatively new radial tracking method was introduced and the backward compatibility with the existing CD-family asked for various provisions. In recent years, the DVD-medium has also developed into a family with recordable

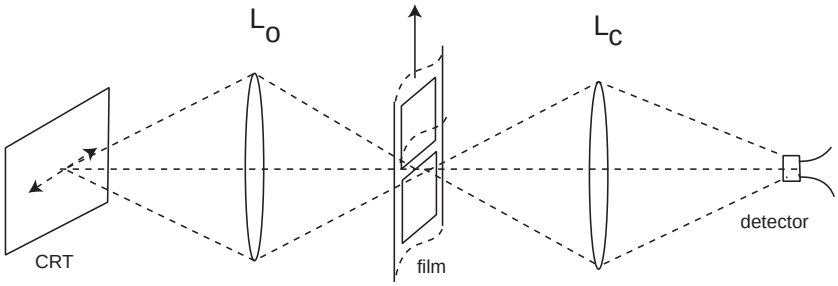

Fig. 2. A sketch of the light paths for the CD, DVD and DVR system and electron microscope photographs with identical magnification of the information pits of the three systems.

and rewritable versions. The main innovation in the DVR system is the use of the deep blue semiconductor laser; combined with a drastically increased value of the numerical aperture of the objective, a forty times higher density with respect to the CD-system is achieved.

Even if the basics of optical disc systems have not changed, an enormous effort in research and development has been needed to adapt the media and the playing units to the drastically increased demands with respect to spatial density, scanning speed and data rate. Another less visible development is the digitisation of the opto-mechanical and electronic control systems in the player. Without the easily programmable digitised servo and control systems, the multi-purpose disc player/recorder that easily accepts a broad range of media would be impossible.

In this chapter we will restrict ourselves to the purely optical aspects of an optical disc player/recorder. More general overviews can be found in the various textbooks mentioned in the introductory section.

3 Overview of the optical principles of the CD- and the DVD-system

In this section we briefly recall the read-out method to extract the stored information from the optical disc. The information is optically coded in terms of a relief pattern on a read-only disc while recordable and rewritable discs employ recording layers that influence both the phase and the amplitude of the incident light. The arrangement of the information is along tracks that have to be followed, both in the recorded

and the ‘blank’ or pre-formatted situation. To this purpose, a bipolar signal is optically derived that can be connected to a servo mechanism to keep the objective head on track during reading or recording. Simultaneously, an error signal is needed to correct for any deviation of the information layer from the optimum focus position.

3.1 Optical read-out of the high-frequency information signal

In Fig.(3), the typical data formats to be read in an optical disc player are shown. The typical track spacing on CD is 1.6 µm and the pit length varies between 0.9 and 3.3 µm with typical increments of 0.3 µm. The data disc in b) has a different

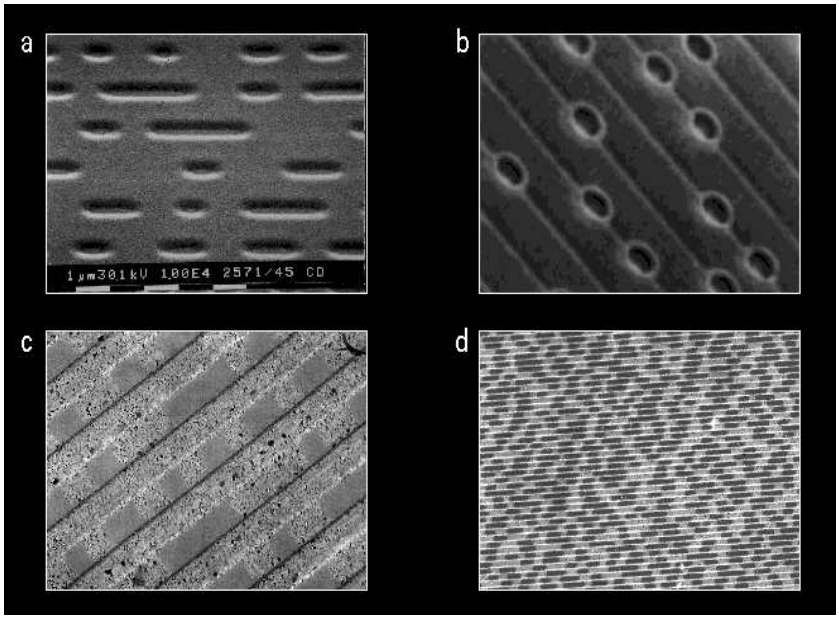

Fig. 3. Typical data formats present on optical storage media.
a) CD-information recorded as a relief pattern in the surface of a disc obtained via injection molding **b**) ‘burned’ data on a digital data disc **c**) amorphous islands (gray) arranged along a pre-formatted track in a micro-crystalline ‘sea’ (phase-change recording) **d**) polarisation microscope photograph of a magneto-optical layer with recorded CD-like information.

coding scheme that initially looked more appropriate for the hole-burning process by avoiding the length modulation. In Fig.(3.c) the CD-ROM format is present on a phase change disc with the low-contrast regions of micro-crystalline and amorphous structure. The disc in d) is not compatible with the standard (or DVD) reading principle because of the polarisation-sensitive read-out.
The reading of the information is done with the aid of the scanning microscope principle that is sketched in Fig.(4). As it is well-known from the literature, the scanning

microscope and the classical microscope are very closely related regarding their resolution and optical bandwidth [19], [20]. In terms of maximum transmitted spatial frequency, the classical microscope attains a value of $2NA/\lambda$ in the case of incoher-

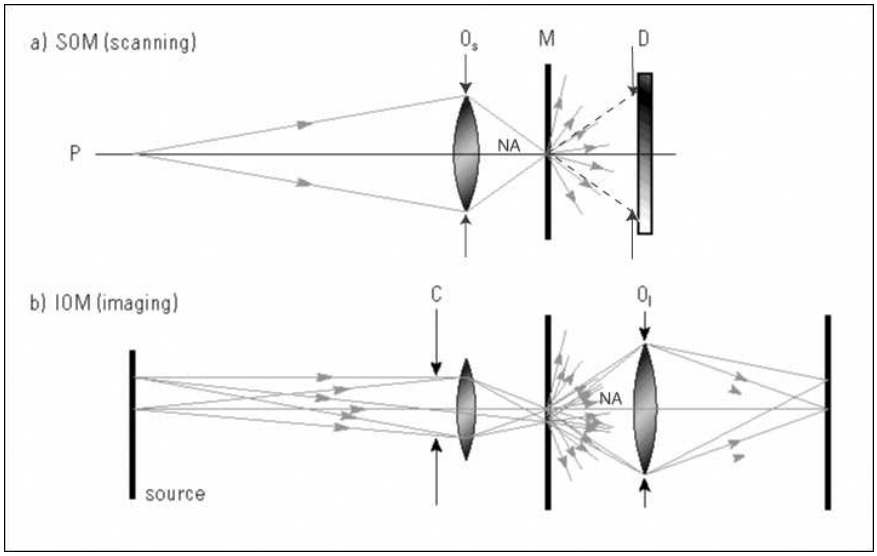

Fig. 4. The scanning optical microscope SOM (**a**) and the 'classical' imaging optical microscope IOM with a full field illumination (**b**). In the scanning microscope, the detector has been schematically drawn and is optically characterised by its angular extent (dashed cone) with respect to the object. In the figure, the angular extent or 'aperture' of the detector has been drawn equal to the numerical aperture NA of the objective O_S, a situation that corresponds to standard reflective read-out of an optical disc.

ent illumination via the condensor system (the numerical aperture of the condensor C then needs to be equal to or larger than the numerical aperture NA of the objective O_I). The scanning microscope achieves the same maximum frequency transfer when the detector D at least captures the same aperture NA as offered by the scanning objective O_S (see heavy dashed lines in Fig.(4.**a**). In the case of a reflective scanning system, the equal aperture of O_S and D is obtained in a straightforward manner because of the double role of O_S as scanning objective and collecting aperture for the detector.

The optical transfer function of a scanning microscope has been visualised in Figs.(5)-(6). The diffraction of a plane wave with propagation angle α by a (one-dimensional) periodic structure (frequency $\nu = 1/p$) leads to a set of distinct plane waves with propagation direction angles α_m given by

$$\sin\alpha_m = \sin\alpha + m\lambda\nu \ , \tag{1}$$

where m is the order number of the diffracted plane wave.

In the case of an impinging focused wave (aperture angle u), the diffracted field

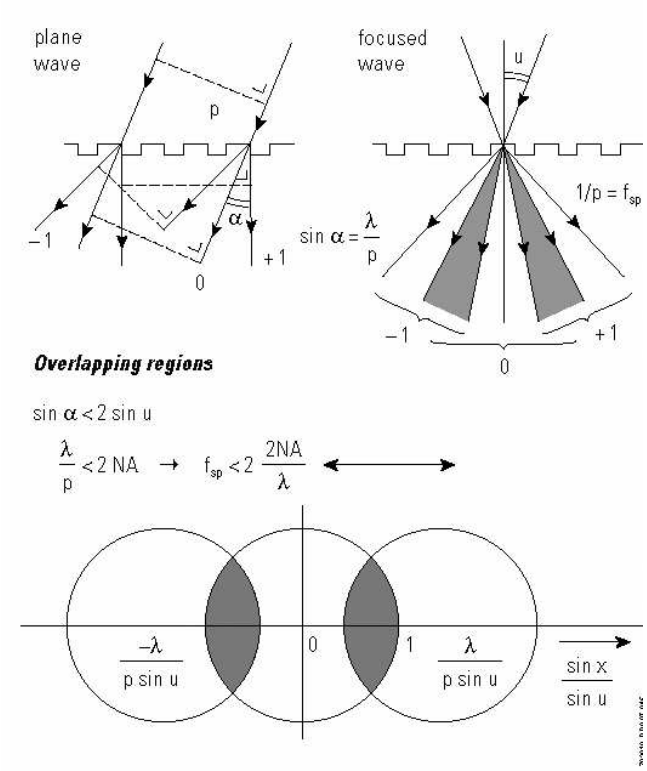

Fig. 5. The diffraction of a plane wave and a focused wave by a one-dimensional periodic structure. The useful region in the far-field for signal detection is given by the regions of overlap of zeroth and first orders.

is a superposition of diffracted spherical waves. The incident spherical wave can be considered to be the superposition of a set of plane waves with an angular distribution ranging form $-u$ to $+u$. Note that in the case of a spherical wave, the diffracted orders can not easily be separated in the far field at a large distance from the periodic structure. This is in particular the case for low spatial frequencies and this is in line with the intuitive feeling that a finely focused wave is not capable of sharply discriminating a large period. As soon as the period becomes comparable with the lateral extent of a focused wave (typically of the order of $\lambda/2NA$), the diffracted spherical waves are fully separated. However, we do not need to fully separate the diffracted spherical orders to draw conclusions about the presence and the position of the periodic structure.

In the regions where e.g. the spherical waves with order number $m = 0$ and $m = 1$ overlap, the two waves are brought to interference and the phase difference between them will depend on the lateral position of the periodic structure with respect to the focal point of the incident spherical wave (note that the phase of a first diffracted or-

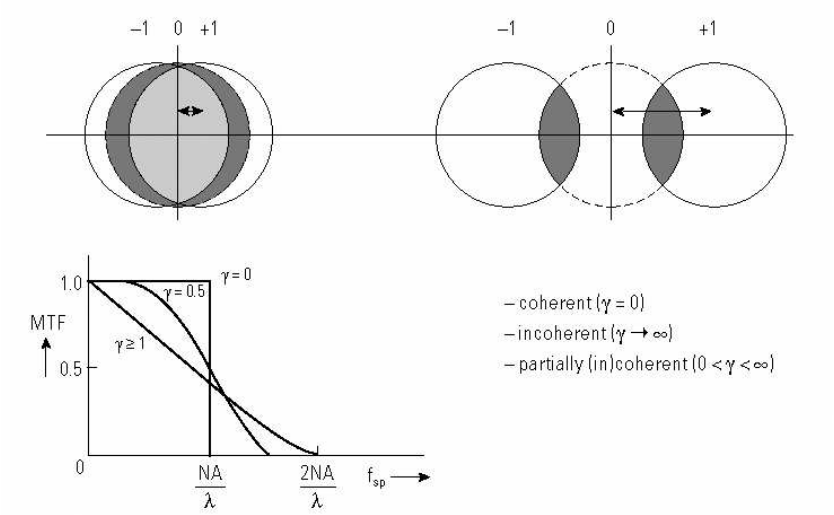

Fig. 6. The normalised optical transfer function as a function of the coherence factor γ, determined by the numerical aperture ratio of detector and scanning objective.

der shows a phase shift ϕ that equals $2\pi\nu\Delta$ if the periodic structure suffers a lateral shift of Δ).

The strength of the interfering signal depends on the size of the overlapping region and on the usefully captured signal by the detector. In the case of a centred detector, the signal depends on its lateral extent and it is normally expressed in terms of the angular beam extent ($NA = \sin u$) of the incident focused beam.

In Fig.(6), the normalised frequency transfer function has been depicted as a function of the commonly called coherence factor $\gamma = NA_D/NA$ (NA_D is the numerical aperture of the detector as defined by geometrical conditions or by a collecting optical system in front of the detector, see [19]). The maximum detectable frequency is $2NA/\lambda$. In this case ($\gamma \geq 1$), the optical system is approaching the fully incoherent limit which means that from the complex amplitude transmission function of an object only the modulus part is detectable. At low values of γ, leading to the limit of 'coherent' read-out, the phase part of an object, even if it is a weakly modulated function, can also be detected in principle. Note that on an optical disc the relief pattern mainly affects the phase part of the optical transmission or reflection function. Because of the strong height modulation with respect to wavelength, the phase modulation can also be detected in the (almost) incoherent mode at γ=1. If the phase modulation switches between 0 and π, the phase modulation even becomes a pure amplitude modulation with extreme values of ± 1. The choice of the value of γ has been fixed at unity from the very beginning of optical storage. In contrast with e.g. optical lithography where a certain minimum transfer value is required and where it can be useful to trade off the maximum transmitted frequency against a better modulation via a choice $\gamma < 1$, the transfer of an optical disc system can

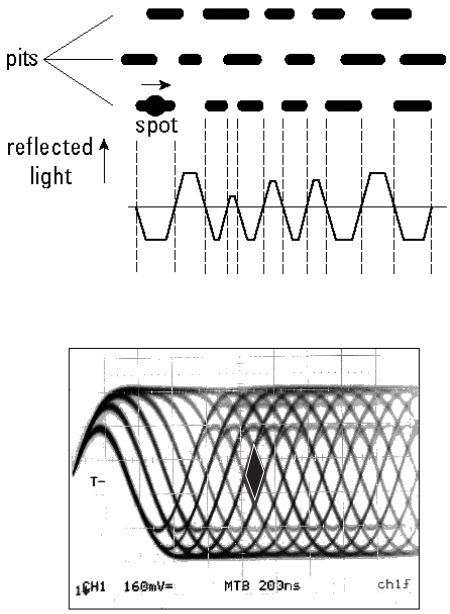

Fig. 7. The pit structure on an optical disc and a schematic drawing of the detected signal that shows a typical maximum slope defined by the spot size λ/NA. An oscilloscope image of a large sequence of signal sections is shown below when the scope is triggered on an arbitrary positive transient in the signal; this signal is commonly called the digital 'eye-pattern' and should preferably not contain transients that are close to the centre of a digital eye (see the drawn lozenge in the figure).

be improved electronically afterwards. This electronic equalisation has proved to be powerful and thus permits to virtually use the entire optical pass-band up to the optical cut-off frequency of $2NA/\lambda$; in practice, frequencies turn out to be detectable up to 90% or 95% of the cut-off frequency. The typical optical frequencies in a CD-system (λ=785 nm, NA=0.45) are 55% of the cut-off frequency for the radial period and approximately 50% for the highest frequency in the direction of the tracks (this is the 3T-3T alternation, the shortest land and pit distances (0.9 µm) on a CD-disc given the clock length of 0.3 µm). Of course, the binary relief pattern in the track of a CD may contain higher frequencies than this highest fundamental frequency. In the DVD-system (λ=650 nm, NA=0.60) the optical pass band is used in a more aggressive way and the typical frequencies are shifted to 75% of the cut-off frequency $2NA/\lambda$.

The typical signal that is obtained from the high-frequency detector in an optical disc player is shown in Fig.(7). A well-opened digital 'eye' is essential for a correct read-out of the stored information.

3.2 Optical error signals for focusing and radial tracking of the information

Focus error signal

Various methods have been proposed for deriving a focus error signal that can be fed to e.g. the coils of a miniature linear electromotor so that an active positioning of the objective with respect to the spinning disc is achieved. In Fig.(8) the

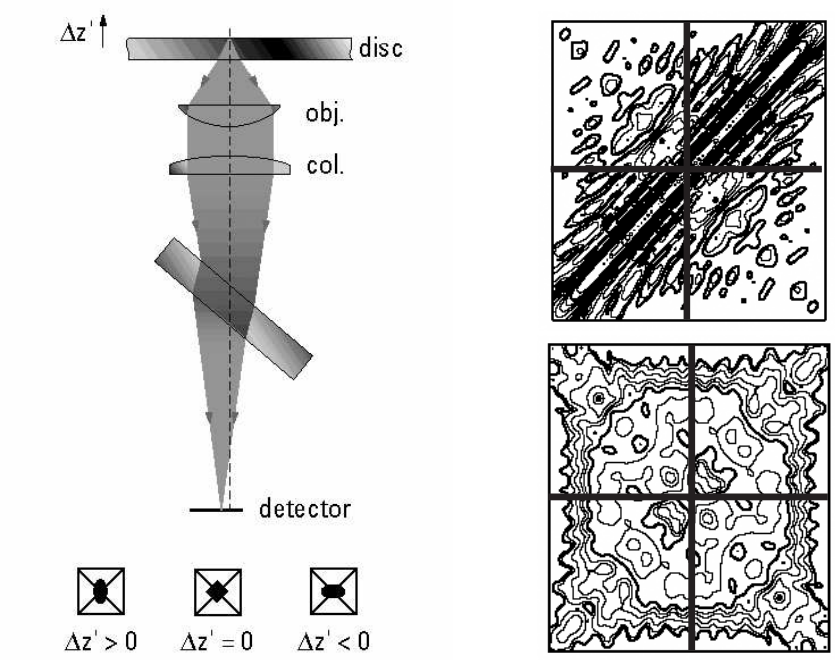

Fig. 8. An optical light path using the astigmatic focusing method by exploiting the astigmatism introduced by a plane parallel plate in the returning converging beam (left). The calculated diffraction patterns corresponding to a focal line and to the best focus position are depicted in the righthand part of the figure, together with the separation lines of the focus quadrant detector (horizontal and vertical in this picture).

commonly used astigmatic method [21] is shown. The light reflected from the disc passes through a relatively thick plane parallel plate (the beam splitter on the way forth from the source to the disc). Depending on the convergence angle of the returning beam and on the thickness of the plate, a certain amount of astigmatism is introduced that generates two focal lines along the optical axis with a 'best' focal region in between them. Projected back onto the information surface and taking into account the axial magnification of the combination of collimator and objective, one aims at a distance between the astigmatic lines of some 10 to 15 μm. The best focus position corresponds to a balanced intensity distribution on the four detectors and should be made to coincide with optimum high-frequency read-out of the optical

disc. In the right part of Fig.(8) a contour picture is shown of the diffraction image in the best focus position and in the case when one of the line images is focused on the detector.
Another candidate for the detection of a focus defect is the Foucault knife-edge test. Although its sensitivity can be higher than that of the astigmatic method, the position tolerances of the knife edge are somewhat critical. A circularly symmetric version of the knife-edge method is the 'ring-toric' lens method [22]. Another option is the 'spot-size' detection method that measures the variation of beam cross-section as a function of defocus. All these methods are capable of delivering a signal with changing polarity as a function of the sign of the focus error. Another important property of an error signal is the maximum distance through which the signal is well above the noise, the so-called effective acquisition range. For both methods, a typical value of 30 to 50 μm can be expected. More refined methods use the detected high-frequency information to monitor the absolute focus-setting and to correct for any drift in the detector balance or the mechanical settings of the light path.

Radial error signal

The methods indicated in Fig.(9) have been devised for the detection of the radial position of the reading or recording light spot with respect to the centre of a track. The first method uses two auxiliary light spots generated by means of a grating (twin-spot method [23]). The light spots have a quarter track spacing off-set with respect to the central light spot. The reflected light belonging to each auxiliary spot is collected in the focal region on separate detectors and the difference in average signal from each detector provides the desired bipolar tracking signal that is zero on-track. An AC-variant of this method with the auxiliary spots uses a small lateral oscillation of the main spot itself. The detection of the phase (either 0 or π) of the oscillation in the detected AC-signal yields the required tracking error signal. The oscillation can be imposed on the scanning spot in the optical player but, preferably, a so-called track wobble is given to the track with a known phase so that no active wobble is needed in the player. Of course, the track wobble should not generate spurious information in the high-frequency signal band nor should it spoil other information (track number, address information) that is stored in an 'undulation' or lateral displacement of the information track.
Both methods described above retrieve the tracking information from a refocused light distribution on one or more detectors that are optically conjugate with the information layers and for this reason they are called near-field methods. A well-known far-field method is the so-called Push-Pull method [21] that uses a duo-detector in subtractive mode for radial tracking (see Fig.(10)). The method is based on the fact that in the presence of a tracking error (the centre of the scanning spot hits e.g. the edges of the information pits) the average propagation direction is slightly bent away from the perpendicular direction. For relatively small values of the phase shift of the light reflected by the pits, the difference signal of the duo-detector shows a change of sign when the scanning spot crosses the information track. If we define the phase depth $\Delta\phi$ of the information pits as the phase shift suffered by the light

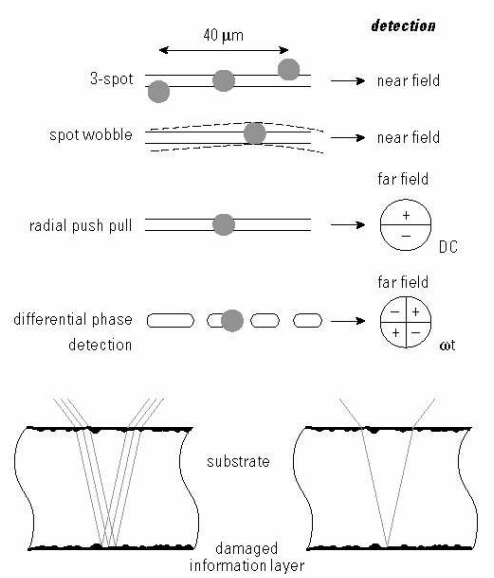

Fig. 9. The various options for deriving a bipolar optical tracking signal. The label 'near field' here means that the detector arrangement is positioned in the image plane of the information layer. In the case of 'far field' detection, the detector is effectively located in the limiting aperture of the objective (or in an image of this diaphragm).

on reflection from the information pits with respect to the light reflected from non-modulated land regions, we observe that the difference signal of the duo-detector reduces to zero if $\Delta\phi$ equals π. The obvious reason is that a phase shift of π between pits and land regions has reduced the information structure to effectively an amplitude structure with no directional phase information available. For this reason, the Push-Pull method is critical with respect to the phase shift induced by the information pits and it also behaves in a rather complicated way once the optical 'effects' have a combined amplitude and phase character (phase-change recording material). A far-field method that, in a first instance, does not suffer from the restrictions on the phase depth $\Delta\phi$ is the differential phase (or time) detection where the phase relations between signal components from a far-field quadrant detector at the instantaneous signal frequency ω are used to derive a tracking signal [24]. A tracking method based on this principle has been standardised for the DVD-system and will be discussed in more detail in a following section.

The possible perturbations of the near-field or far-field methods is correlated with the conjugate position of the respective detectors. In Fig.(9), lower part, the near-field methods should be sensitive to damage of the information layer while the far-field based methods should be more sensitive to damage or scratching of the upper surface (non-info) of the disc. Because of the limited bandwidth of the tracking error signals, their (average) information is collected from a relatively large disc area

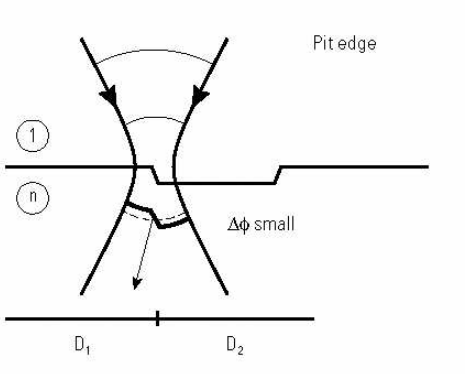

Fig. 10. The phase change introduced by a pit edge to the focused wave (here shown in transmission). The average propagation direction of the transmitted beam is slightly tilted causing unbalance between the detectors D_1 and D_2.

(a stripe on the info-layer, an elongated circle on the non-info surface). Because of the spatial averaging effect, no substantial difference of preference is found for the near- or far-field methods.

3.3 Examples of light paths

The various methods for deriving the high-frequency signal and the optical error signals have to be incorporated in a light path that has wide tolerances both during manufacturing and product life (optical damage, thermal and mechanical drifts). In Fig.(11), lefthand side, a classical light path for CD is shown with the light from the laser diode source coupled in via a semi-transparent plate towards the collimator-objective combination. The reflected light is transmitted through the beam splitter towards a double-wedge prism that separates the far field region in order to accommodate for the detection of a Push-Pull radial tracking signal. The double wedge also serves as 'knife-edge' for each beam half thus producing two complementary knife-edge images for focus detection. This light path is made from simple, easily available components. The tightest tolerance regarding temperature and mechanical drift is found in the relative position of source and detector with respect to the beam splitter.
In the righthand figure, the beam splitting action is obtained with the aid of a holographic element. The relatively thick substrate carries a grating on the lower surface for generating the three spots needed for radial tracking. The upper surface contains a more complicated structure, the holographic or diffractive element. At the way forth to the disc, the three beams generated by the grating continue as the zeroth order of the holographic structure and they are captured by the collimator and objective. Consequently, three light spots are focused on the disc. On the way back, the light diffracted by the hologram propagates towards the detector region and a far-field splitting is obtained by sending one half circle of the beam cross-section to

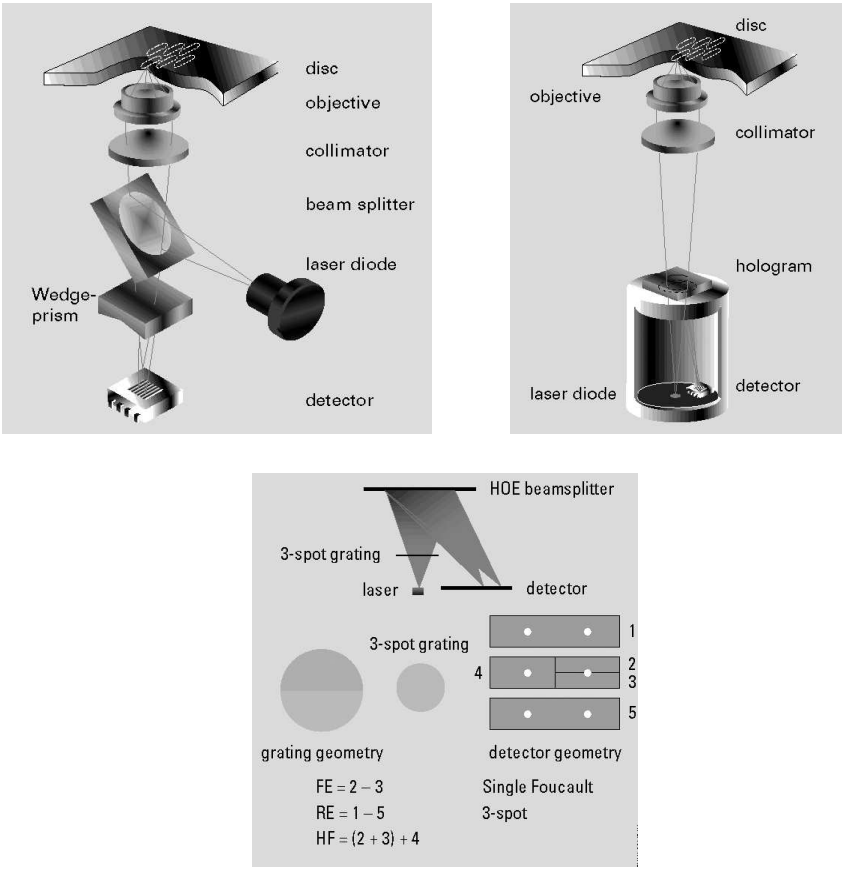

Fig. 11. An optical light path using separate optical components for achieving the beam separation and the generation of the focus and radial error signals (left). In the righthand picture, the splitting function and the generation of the error signals is obtained by means of holographic or diffractive optical element. In the lower picture, the various beams are shown that propagate towards the detector after traversal of the holographic beam splitter HOE.

the detectors 2 and 3 and the other half to detector 4. The auxiliary spots produced by the grating are directed towards the detectors 1 and 5. A focus error signal is obtained by taking the difference between detectors 2 and 3 while the Push-Pull radial tracking error signal can be obtained by the signal combination (2+3)-4. Moreover, an extra twin-spot radial error signal is obtained from the detector difference signal 1-5 (the radial offset of each spot is one quarter of a pitch). The high frequency signal is derived from the sum of the detectors 2, 3 and 4. The attractiveness of the hologram solution is its mechanical and environmental stability. The directions of the diffracted orders are rather insensitive to hologram tilt and displacement and the source and detector are mounted on one single substrate. A wavelength change will

shift the diffracted spots along the dividing line between detectors 2 and 3 but does not cause an offset.

4 Radial tracking for DVD

Some disadvantages of the radial tracking methods had become manifest while using them for the CD-system. The twin-spot method requires an extra grating with a rather delicate angular orientation and, in the recording mode, precious optical power (some 20%) is spread out to the auxiliary twin spots. The Push-Pull method is sensitive to the phase depth $\Delta\phi$ of the information and does not work in the presence of pure amplitude information patterns. The spot or track wobble method, due

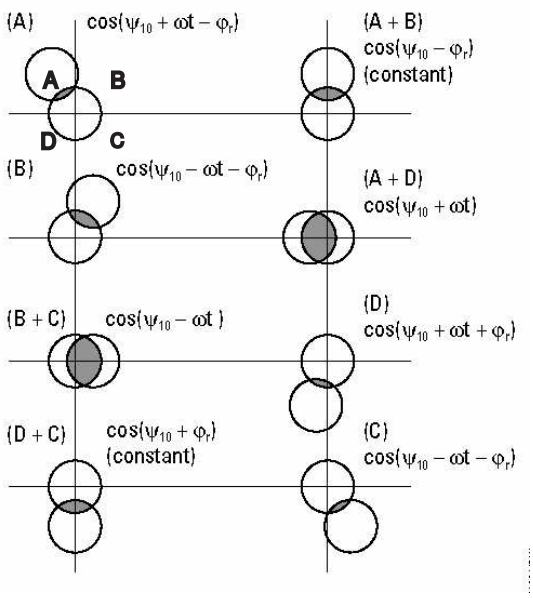

Fig. 12. Overlapping regions in the far-field quadrants A, B, C and D with typical modulation terms in the DC-domain and in the high-frequency domain (frequency ω). In each of the eight examples shown, the letters are printed (between brackets) of the quadrants in the far field where a certain harmonic modulation term of the signal is detected. The dependence on the disc translation is given by the term ωt and a radial displacement by the phase ϕ_r. If no translation term ωt is present, the signal only shows variations of average instensity (denoted by 'constant') due to radial position. The tangential direction is chosen horizontally in the figure.

to its very nature, consumes precious space in the radial direction and is likely to increase the cross-talk between tracks when compared to the other methods at equal

density. For this reason, the Differential Phase Detection (DPD) method [24], originally devised for the analog video disc system, has been revisited to adapt it to the digital signals on the high-density DVD disc. It is now commonly called the Differential Time Detection (DTD) method

4.1 A diffraction model for the DPD and DTD tracking signal

An easy picture for the understanding of the differential phase (or time) detection method is obtained when the information in the tangential direction is replaced by a single frequency, so that, together with the radial periodicity, a purely two-dimensional grating is present on the disc. In Fig.(12) the various quasi-DC and AC signals are depicted that are present in the far-field on a quadrant detector (quadrants A, B, C and D). Each modulation term depends on the diffraction direction.

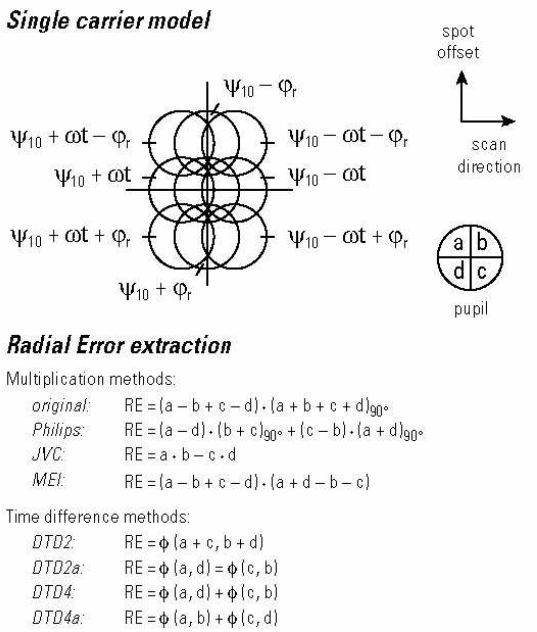

Fig. 13. Various methods for the extraction of a Differential Time radial error signal using the signals present in the far-field quadrants.

Radially diffracted orders show a (varying) intensity level in an overlapping region with the zeroth order that is proportional to $\cos(\psi_{10} \pm \phi_r)$. The reference phase ψ_{10} between the first and zeroth orders is determined by the disc structure; it varies from $\pi/2$ for very shallow structures to π if the phase depth $\Delta\phi$ of the optical effects attains the value of π. The quantity $\phi_r = 2\pi r/q$ is proportional to the tracking error

r (q is the track pitch). The sign of ϕ_r depending on the order number ± 1 of the diffracted order. In the same way, the overlapping regions in the tangential direction show an intensity variation according to $\cos(\psi_{10} \pm \omega t)$, caused by the scanning with uniform speed of the tangentially stored periodic pattern. Obliquely diffracted orders show a mixed phase contribution and are carrying the interesting information for deriving the tracking error.
Collecting the interference terms on the various quadrants, we obtain

$$\begin{aligned} S_A(t,\phi_r) &= \cos(\omega t + \psi) + \alpha\cos(\omega t - \phi_r + \psi) \\ S_B(t,\phi_r) &= \cos(\omega t - \psi) + \alpha\cos(\omega t + \phi_r - \psi) \\ S_C(t,\phi_r) &= \cos(\omega t - \psi) + \alpha\cos(\omega t - \phi_r - \psi) \\ S_D(t,\phi_r) &= \cos(\omega t + \psi) + \alpha\cos(\omega t + \phi_r + \psi)\ , \end{aligned} \tag{2}$$

where all possible reference phase angles between zeroth and first diffracted orders, for reasons of simplicity, have been replaced by a single phase angle ψ and where α is a factor less than unity that accounts for the relatively small contribution of the diagonal orders with respect to the tangential orders.
With this approximation, the various high-frequency signals are now given by:

$$\begin{aligned} S_{CA}(t,\phi_r) &\propto +\{1+\alpha\cos\phi_r\}\cos\psi\cos(\omega t) \\ S_{tPP}(t,\phi_r) &\propto +\{1+\alpha\cos\phi_r\}\sin\psi\sin(\omega t) \\ S_{rPP}(t,\phi_r) &\propto +\alpha\sin\phi_r\sin\psi\cos(\omega t) \\ S_{dPP}(t,\phi_r) &\propto -\alpha\sin\phi_r\cos\psi\sin(\omega t)\ \ . \end{aligned} \tag{3}$$

The subscripts here denote the signals according to the detector schemes Central Aperture (CA=$A+B+C+D$), tangential Push-Pull (tPP=$A-B-C+D$), radial Push-Pull (rPP=$A+B-C-D$) and diagonal Push-Pull (dPP=$A-B+C-D$). For the derivation of a tracking signal, a reference signal with a well-defined time-dependent phase is needed for comparison with a signal whose phase depends on the tracking error. As a reference signal we can choose the Central Aperture (CA) signal or the tangential Push-Pull signal. If we take the CA-signal and multiply it with the dPP-signal after a phase shift of the latter over $\pi/2$ we obtain after low-pass filtering:

$$S_1(\phi_r) \propto -\alpha\cos^2\psi\{\sin\phi_r + \frac{1}{2}\alpha\sin 2\phi_r\}\ \ . \tag{4}$$

The multiplication of the tangential PP-signal and the diagonal PP signal directly yields after low-pass filtering:

$$S_2(\phi_r) \propto -\alpha\sin\psi\cos\psi\{\sin\phi_r + \frac{1}{2}\alpha\sin 2\phi_r\}\ \ . \tag{5}$$

Inspection of the signals $S_1(\phi_r)$ and $S_2(\phi_r)$ shows that they provide us with the required bipolar signal of the tracking error ϕ_r. The factor in front, containing ψ, depends on the phase depth $\Delta\phi$ and makes $S_2(\phi_r)$ less apt for discs with relatively

deep optical structures while $S_1(\phi_r)$ performs optimum in that case.
Other possible combinations are obtained when combining the high-frequency radial Push-Pull signal with either the CA- or the tPP-signal. One easily deduces that the following signals result after multiplication, phase-shifting and low-pass filtering of, respectively, rPP and CA ($S_3(\phi_r)$ and rPP and tPP Radial ($S_4(\phi_r)$:

$$S_3(\phi_r) \propto \alpha \sin\psi \cos\psi \{\sin\phi_r + \frac{1}{2}\alpha \sin 2\phi_r\} \quad . \tag{6}$$

$$S_4(\phi_r) \propto \alpha \sin^2\psi \{\sin\phi_r + \frac{1}{2}\alpha \sin 2\phi_r\} \quad . \tag{7}$$

The ψ-dependence of especially $S_4(\phi_r)$ makes it unsuitable for standard optical discs but especially favourable for extra shallow disc structures.

4.2 The influence of detector misalignment on the tracking signal

The single-carrier model can be equally used for studying the effects of misalignment of the quadrant detector with respect to the projected far-field pattern, the so-called beam-landing effect. This effect is produced during quick access of a remote information track. The sledge (coarse movement) and the scanning objective (precise positioning) are actuated together and this action produces a lateral displacement of the objective with respect to the fixed collimator and detector that can amount to some 10% of the beam foot-print on the detector. As a consequence, the whole diffraction pattern is shifted on the detector in the radial direction (see the arrow 'spot offset' in Fig.(13)). Taking into account such a mechanically induced offset with a size of ϵ (normalised with respect to the half-diameter of the beam foot-print), we write the signals corresponding to the interference terms in the overlapping regions of the far field as

$$\begin{aligned}
S_A(t,\phi_r) &= (1+\epsilon)\cos(\omega t + \psi) + \alpha\cos(\omega t - \phi_r + \psi) \\
S_B(t,\phi_r) &= (1+\epsilon)\cos(\omega t - \psi) + \alpha\cos(\omega t + \phi_r - \psi) \\
S_C(t,\phi_r) &= (1-\epsilon)\cos(\omega t - \psi) + \alpha\cos(\omega t - \phi_r - \psi) \\
S_D(t,\phi_r) &= (1-\epsilon)\cos(\omega t + \psi) + \alpha\cos(\omega t + \phi_r + \psi) \quad .
\end{aligned} \tag{8}$$

The beam-landing effect has only been applied to the purely tangential diffraction orders; the diagonal orders are rather far away from the horizontal dividing line of the quadrants.
Two of the four basic signals derived from the quadrants A, B, C and D depend on the radial misalignment and they become:

$$\begin{aligned}
S_{rPP}(t,\phi_r) &\propto \alpha \sin\phi_r \sin\psi \cos(\omega t) - \epsilon \cos\psi \cos(\omega t) \\
S_{dPP}(t,\phi_r) &\propto -\alpha \sin\phi_r \cos\psi \sin(\omega t) - \epsilon \sin\psi \sin(\omega t) \quad .
\end{aligned} \tag{9}$$

The four possible DPD-signals now become

$$S_1(\phi_r) \propto -\alpha\cos^2\psi\{\sin\phi_r + \frac{1}{2}\alpha\sin 2\phi_r\} - \epsilon\sin\psi\cos\psi(1+\alpha\cos\phi_r)$$
$$S_2(\phi_r) \propto -\alpha\sin\psi\cos\psi\{\sin\phi_r + \frac{1}{2}\alpha\sin 2\phi_r\} - \epsilon\sin^2\psi(1+\alpha\cos\phi_r)$$
$$S_3(\phi_r) \propto \alpha\sin\psi\cos\psi\{\sin\phi_r + \frac{1}{2}\alpha\sin 2\phi_r\} - \epsilon\cos^2\psi(1+\alpha\cos\phi_r)$$
$$S_4(\phi_r) \propto \alpha\sin^2\psi\{\sin\phi_r + \frac{1}{2}\alpha\sin 2\phi_r\} - \epsilon\sin\psi\cos\psi(1+\alpha\cos\phi_r) \quad . \tag{10}$$

The signals S_1 and S_4 show a beam landing sensitivity that is maximum for the $\lambda/4$ pit depth ($\psi \approx 3\pi/4$); for deep pits (optical depth is $\lambda/2$) or for amplitude structures ($\psi = \pi$) the beam landing sensitivity is zero.
An interesting combination of the signals S_1 and S_4 is given by

$$S_5(\phi_r) = S_1(\phi_r) - S_4(\phi_r) \propto \{\sin\phi_r + \frac{1}{2}\alpha\sin 2\phi_r\} \quad . \tag{11}$$

The detection scheme can be written as follows:

$$S_5(\phi_r) = [B-C][A+D]^{90^o} - [A-D][B+C]^{90^o} \quad , \tag{12}$$

where the index 90^o implies that the 90 degrees phase shifted version of the signal has to be taken. The signal S_5 is fully independent of pit depth and does not suffer at all from beam landing off-set.
Another possible linear combination is given by

$$\begin{aligned} S_6(\phi_r) &= S_2(\phi_r) - (1+s)S_3(\phi_r) \\ &\propto -(1+s/2)\sin\psi\cos\psi\{\sin\phi_r + \frac{1}{2}\alpha\sin 2\phi_r\} \\ &\quad +\epsilon/2\,[\sin^2\psi - (1+s)\cos^2\psi]\{1+\alpha\cos\phi_r\} \quad . \end{aligned} \tag{13}$$

The beam landing influence on this signal is reduced to zero by the condition

$$s = \tan^2\psi - 1 \quad . \tag{14}$$

The detection scheme for S_6 is

$$S_6(\phi_r) = [CD - AB] + (s/4)[(C+D)^2 - (A+B)^2] \quad . \tag{15}$$

In the particular case that $\psi = 3\pi/4$, the signal is free from beam landing effects when using its simplest form $[CD - AB]$; the corresponding tracking signal is obtained after low-pass filtering of this high-frequency signal product.

4.3 The DTD tracking signal for the DVD-system

Shifting from analog to digital signals with their associated clock reference signal, it is evident that the measured phase shifts will now be replaced by time differences

between the clock signals associated with specific digital signals derived from the quadrant detector.
The principle of differential time detection can be explained with the aid of Fig.(14). The focused light spot (centre in M) is scanning an information pit slightly off-track (distance v_0) in the direction of the arrow. The leading and the trailing edge of the

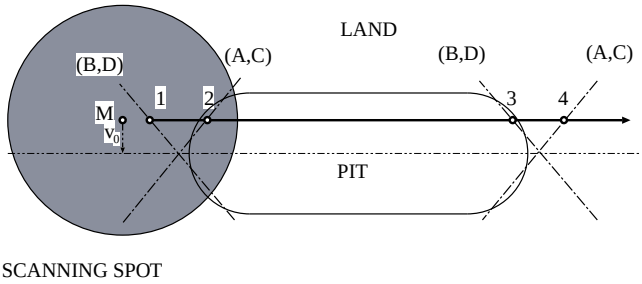

Fig. 14. A schematically drawn focused spot (gray area) is scanning an optical pit with a tracking error v_0. The positions 1 and 3 mark the front and end of the pit as they are detected by the detector pair (B, D); the positions 2 and 4 correspond with the beginning and end of the pit when detected by the diagonal detector pair (A, C).

pit induce diffraction of the light perpendicularly to the hatched lines (A, C) and (B, D). When the centre M of the scanning spot is in the position 1, the detector pair (B, D) perceives the leading edge of the pit and the intensity on this detector pair will go down. A short time later, the detector pair (A, C) will measure the passage of the leading edge because the intensity goes down at the position 2. Once the scanning spot has advanced towards the pit end, the trailing edge will be detected at the positions 3 and 4 by a rise in intensity on respectively the detector pairs (B, D) and (A, C). By measuring the time difference between the information signal on the diagonal detector pairs, a bipolar function of the tracking error v_0 can be obtained. Further away from the edges, in the longer pit sections and land regions, the summed intensities on the two quadrant pairs are virtually equal and there is no contribution to the DTD-signal.
The standard DTD-signal is derived by first summing the intensities on a diagonal pair of detector and then detecting the time shift by electronic means. It is also possible to compare the phase difference between individual quadrants and we conclude that the following set of detection schemes is possible:

- $\tau_{(A+C)}$ - $\tau_{(B+D)}$
  The time difference between the sum signal from quadrants $A + C$ and the sum signal from quadrants $B + D$ is taken. This is the standard DTD-method based on a diagonal difference signal and the denomination is $DTD2$-signal.
- (τ_B - τ_C) + (τ_D - τ_A)
  This difference signal requires four independent high-frequency detectors; on track, each difference signal $(D - A)$ and $(B - C)$ is identical zero. The com-

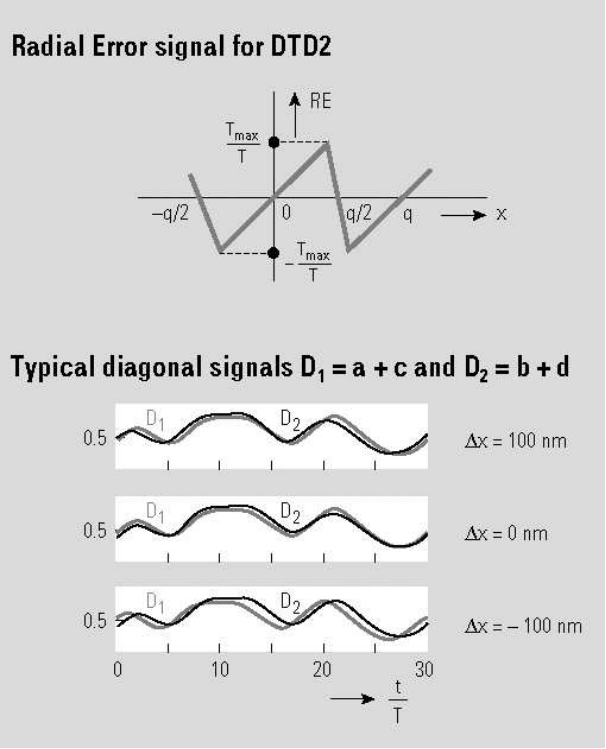

Fig. 15. The mutual shift in time of the diagonal quadrant signals D_1 and D_2 as a function of the tracking error and the resulting radial error signal, schematically drawn in the upper part of the figure as a function of the radial off-track position with periodicity q.

mon denomination is $DTD4$-signal

- (τ_B - τ_A) + (τ_D - τ_C)
  This signal is comparable to the preceding one; the time differences $B - A$ and $D - C$ are not zero on track but depend, among others, on the optical pit depth.

4.4 The DTD2 and the DTD4 signal in the presence of defocus

In the case of the standard analog DPD-signal, we have seen that the phase reference value ψ between the zeroth order and the first order diffracted light played an important role via a $\cos^2\psi$-dependence. It is a well-known fact that the effective phase difference between zeroth and first order light is affected by a defocusing. Zernike's discovery of phase contrast was triggered by his observation that a phase grating image showed beautiful contrast in a defocused setting. This focus-dependence of ψ means that the the amplitude of the standard DPD-signal but also the $DTD2$-signal will vary when defocus is present. If $\psi = \pi$, the signal amplitude

is symmetric around focus and the maximum is found in the optimum focus setting. If $\psi < \pi$, the tracking signal will show a maximum value in a defocused situation and can even become zero in focus. Special DPD-signals ($S_5(\phi_r)$ and $S_6(\phi_r)$) remained unaffected by the value of ψ and they also show a behaviour that, in a first instance, is independent of defocus. It turns out that for the DTD4-signal, where the time difference is evaluated independently for each quadrant detector, a comparable robustness holds with respect to defocus. In Fig.(16) the $DTD2$ and $DTD4$

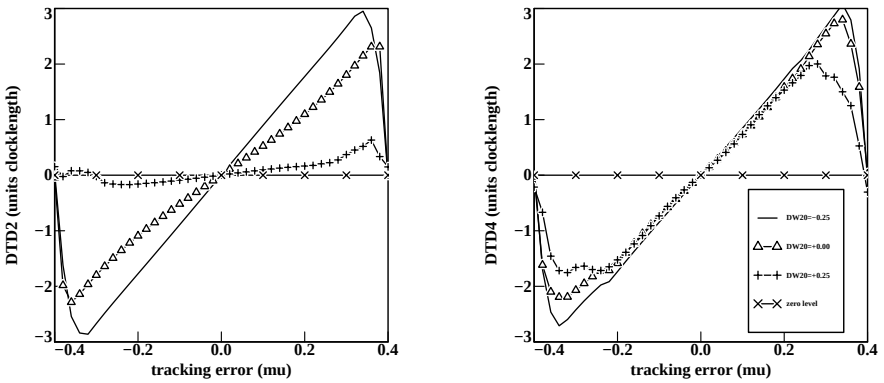

Fig. 16. The radial tracking signals $DTD2$ and $DTD4$ with different amounts of defocus. The defocus parameter $DW20$ is expressed in wavefront aberration at the pupil rim in units of the wavelength λ of the light. A value of 0.25 corresponds to a defocusing Δz_s of one focal depth (typically 0.8 μm for a DVD-system). The tracking error signal has been numerically computed using a long sequence of pits obeying the standard EFM (Eight to Fourteen Modulation) scheme [25]. The disk layout and the reading conditions are those encountered in the DVD-system but the depth of the information pits has been reduced to 60 nm ($\psi \approx 3\pi/4$) instead of the common 100 nm.
drawn line: Δz_s=-0.8 µm, triangles: Δz_s=0, crosses: Δz_s=+0.8 µm.

signals have been depicted for a rather shallow disk pattern with DVD-density. It is clear from the figure that there is a pronounced asymmetry around focus for the $DTD2$-signal while the $DTD4$-signal virtually remains unaffected by defocus; in particular the slope at the central zero set-point of the servo system remains unaltered on defocusing.
We conclude by reminding that the introduction of the DTD-tracking method in the CD-system was hampered by a signal instability that appeared when neighbouring track portions showed comparable information with a fixed mutual position over a distance that was within the bandwidth of the radial servo system. An example of such a problematic situation was found in the fixed digital pattern used to represent absolute silence on a CD audio disc. In the DVD-system, this fixed pattern problem

has been avoided by an appropriate extra coding step that suppresses the prolonged appearance of fixed patterns.

5 Compatibility issues for the DVD- and the CD-system

The compatibility issues between the two standards CD and DVD are caused by the change in substrate thickness, the change in wavelength and, to a lesser degree, the change in numerical aperture. The combination of smaller wavelength and higher NA would lead to a density increase by a factor of 2.5. Further gain in density has been obtained by the higher frequency choice in the DVD-system with respect to the available optical pass-band; this has led to a total factor of five. The final factor of 7.2 has been achieved by more efficient modulation schemes and error correction techniques.
In Fig.(17) the specifications of both systems have been tabulated and we observe that the substrate thickness has been halved for DVD. The reason for this reduction in thickness was the specification on disc tilt during playback. The higher NA and shorter wavelength of the DVD-system causes a two times larger comatic aberration of the read-out spot at equal tilt angle. The amount of coma being linearly dependent on the substrate thickness, a reduction of thickness was required. The 600 µm compromise between tolerances and mechanical stability was solved by bonding two substrates together, leading to a doubled capacity. As it appears from Fig.(18), the doubled capacity can be further increased by introducing in both substrates a buried information layer that is approximately semi-transparent. The distance between the buried layer and the fully reflecting information layer is specified in a margin of 40 to 70 µm to avoid a too large value of spherical aberration when reading the layers at an NA of 0.60 in red light (λ=650 nm).

5.1 The substrate-induced spherical aberration

In a DVD player, the most critical task is the reading of the high-density DVD disc with an NA value of 0.60 and with light of a wavelength of typically 650 nm. Although a CD or CD-ROM disc could also be read with the same wavelength, this does not necessarily apply to a CD-R or a CD-RW disc. The correct optical contrast between recorded optical effects and the land regions is only guaranteed at the original CD-wavelength of 780 nm. The reading of a CD-type disc thus first requires an extra source and, secondly, a solution needs to be found for the reading through a substrate that is 600 µm thicker than prescribed in the specification for the DVD-objective. The thicker substrate causes spherical aberration in the CD-reading spot due to the fact that the marginal rays are more refracted than the paraxial rays, thus enlarging the focal distribution in the axial direction away from the objective. Simultaneously, a serious lateral smearing out of the light intensity distribution is observed and the correct reading of the CD-information is prohibited by this aberration phenomenon. In Fig.(19), images of intensity distributions at various focus settings are shown for a focused wave that suffers from spherical aberration. The

CD and DVD Specification

	CD	DVD
Disc diameter	120 mm	120 mm
Disc thickness	1.2 mm	1.2 mm
Disc structure	Single substrate	Two bonded 0.6 mm substrates
Laser wavelength	780 nm (infrared)	650 and 635 nm(red)
Numerical aperture	0.45	0.60
Track pitch	1.6 µm	0.74 µm
Shortest pit/land length	0.83 µm	0.4 µm
Reference speed	1.2 m/sec CLV	4.0 m/sec CLV
Data layers	1	1 or 2
Data capacity	Approx. 680 megabytes	Single layer: 4.7 gigabytes
		Dual layer: 8.5 gigabytes
Reference user data rate	Mode 1: 153.6 kilobytes/sec Mode 2: 176.4 kilobytes/sec	1,108 kilobytes/sec 1.1 Mbytes/sec

Video format

	Video CD	DVD-video
Video data rate	1.44 megabits/sec (video,audio)	1 to 10 megabits/sec. variable (video,audio, subtitles)
Video compression	MPEG1	MPEG2
Sound tracks	2 Channel-MPEG	Mandatory (NTSC): 2-channel linear PCM; 2-channel/5.1-channel AC-3 Optional: up to 8 streams of data available
Subtitles	Open caption only	Up to 32 languages

Fig. 17. Table with the most relevant specifications of the CD- and the DVD-system.

picture at the height of the chain-dotted line is found at the axial position corresponding to the paraxial focus; approximately two pictures up, the marginal focus position is found where the aberrated marginal rays cut the optical axis. The amount of wavefront spherical aberration W_S, written as a power series of the numerical aperture (NA=$\sin\alpha$), and normalised with respect to the wavelength of the light is given by

$$W_S = \frac{d}{\lambda}\left\{\left(\frac{n^2-1}{8n^3}\right)\sin^4\alpha + \left(\frac{n^4-1}{16n^5}\right)\sin^6\alpha +\right.$$

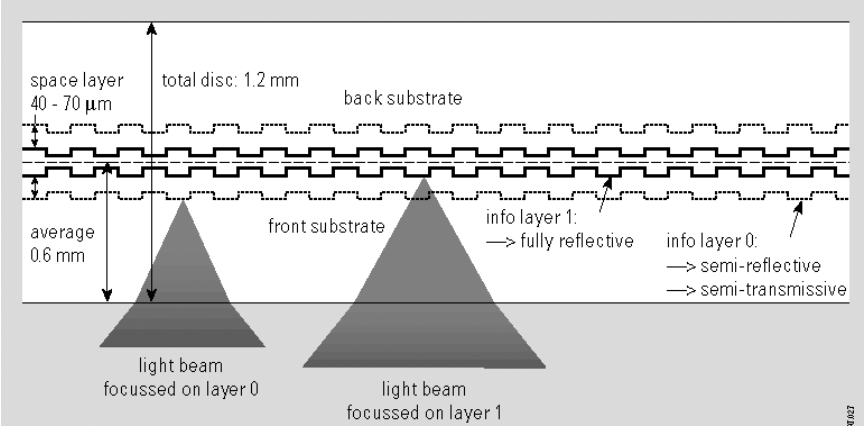

Fig. 18. Cross-section of a double-sided DVD-disc with an extra semi-transparent information layer (dotted) on each side.

$$\left(\frac{5(n^6-1)}{128n^7}\right)\sin^8\alpha+\cdots\Big\}\ , \tag{16}$$

where n is the refractive index of the disc substrate material and d the thickness deviation.

If we use the numbers λ=650 nm, d=600 µm, n=1.586 (refractive index of polycarbonate) and $\sin\alpha$=0.60, the numerical values of the three factors in the power series of $\sin\alpha$ in Eq.(16) amount to, respectively, 5.68, 1.43 and 0.36 in units of wavelengths. At a first sight, knowing that optical disc read-out has to be done in diffraction-limited conditions, these values are far too large. In Fig.(20) we have depicted the behaviour of the central intensity along the axis and in Fig.(21) we present calculated intensity distributions at axial positions between the paraxial and the 'best' focus (halfway the paraxial and the marginal focus). From Fig.(21), it is obvious that the reading of a CD-disc is impossible close to the 'best' focus setting (e.g. using the light distribution from the second row, last picture) . The reading spot is very much degraded and almost all energy has drifted towards strong diffraction rings; at read-out, the diffraction rings will yield an unacceptable intersymbol interference and the digital eye pattern fully disappears. However, the asymmetry with respect to the 'best' focus position shows a peculiar effect: towards the paraxial region, the strong diffraction rings disappear and a central lobe is observed superimposed on a reduced intensity background. The central lobe is much broader than in the best focus but could have an acceptable size for reading a low-density CD-disc when the background intensity is somehow made inoffensive by the read-out method. The distribution that most resembles a standard reading spot (appropriate half width, low level of side lobes) is found around the third position of Fig.(21). The focus setting corresponds to an offset from the paraxial focus of some 6 to 8 focal depths (one focal depth $\approx$ 0.8 µm) and yields an intensity distribution that should be adequate for CD read-out. In Fig.(20), this interesting axial position is

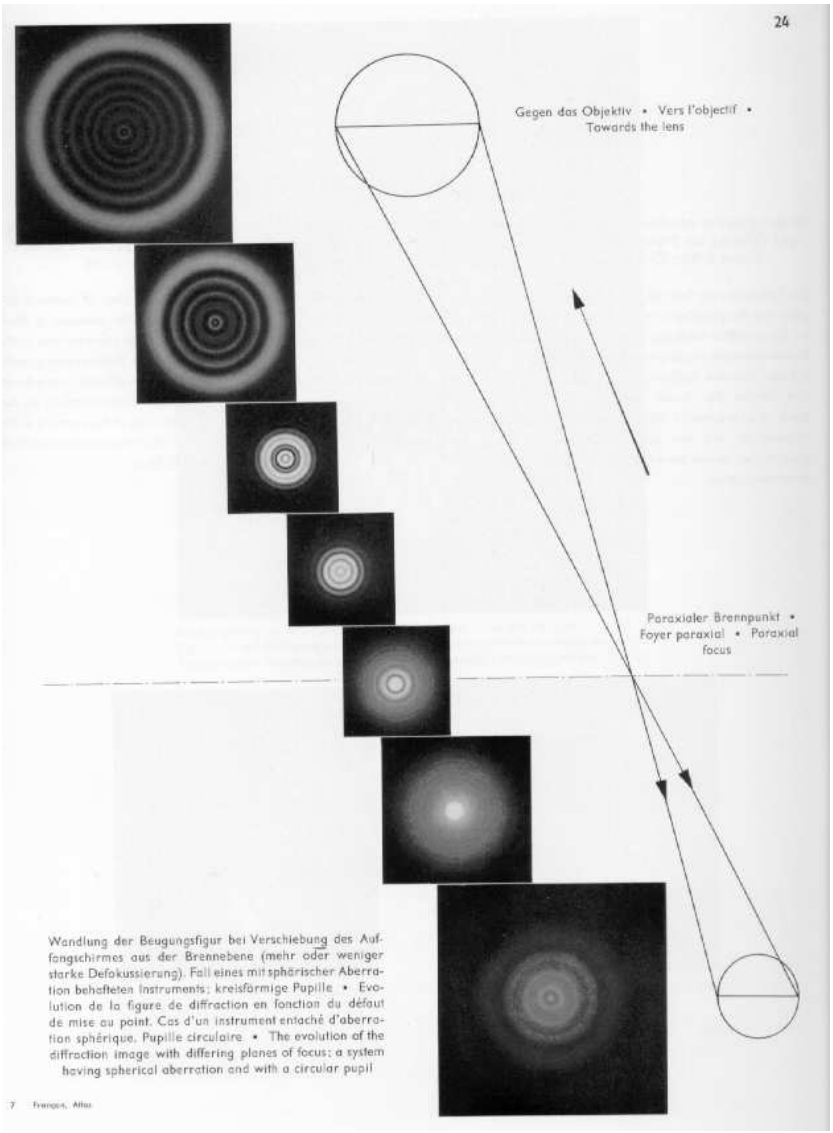

Fig. 19. Intensity profiles at different positions along the optical axis for a pencil suffering from spherical aberration (from: M. Cagnet, M. Françon and J.-C. Thrierr, Atlas of optical phenomena, Springer-Verlag, Berlin, 1962). In the case of a DVD-player reading a CD-disc, the extra substrate thickness introduces spherical aberration with a sign such that the lower part of the figure is closest to the objective and the light travels in the direction given by the large arrow in the upper part of the figure.

approximately halfway up-hill the first maximum.

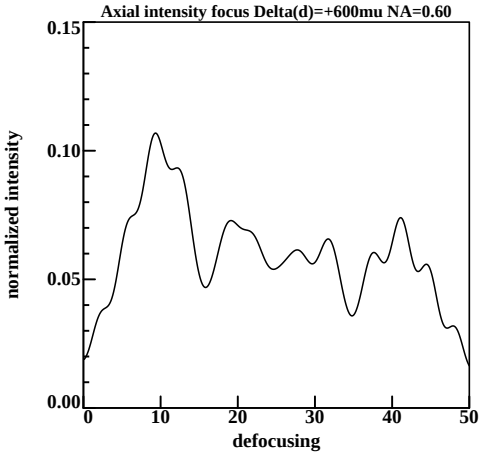

Fig. 20. A plot of the axial intensity of a pencil suffering from spherical aberration due to a substrate thickness increase of 600 μm (NA=0.60, λ=650nm). The paraxial focus is at z=0, the best focus at z=24 μm and the marginal focus at z=48 μm.

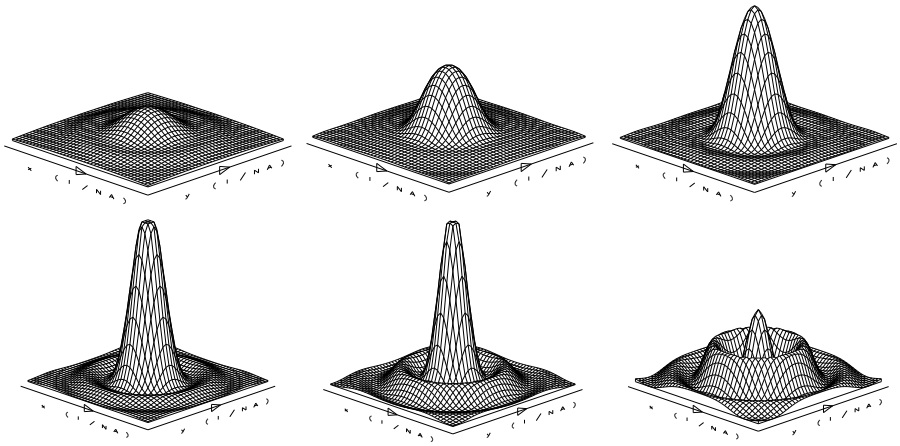

Fig. 21. 3D-plots of the intensity distribution (from left to right) going away from the paraxial focus towards the best focus in axial steps of 3 μm (one focal depth equals 0.81 μm (NA=0.60, λ=650nm). The maximum intensity in the figures is about 0.10 with respect to 1.0 for the diffraction-limited case. The length and width of the square plot area are $4\lambda/NA$ (4.33 μm). The upper right figure represents the optimum focus setting for reading a CD-disc.

Given the low peak intensity, an important amount of low-intensity aberrated stray light can be expected around the central peak and this background should be prevented from reaching the detector. This can be done by introducing an aperture in the light path towards the detector, or, alternatively, by using a detector of finite size; the strongly aberrated rays, after double passage through the mismatched substrate, will have a large transverse aberration in the image plane and miss the quadrant de-

tector. The effect on the calculated digital eye pattern of the effective reduction of the detector aperture NA_D is illustrated in Fig.(22). These graphs show the result-

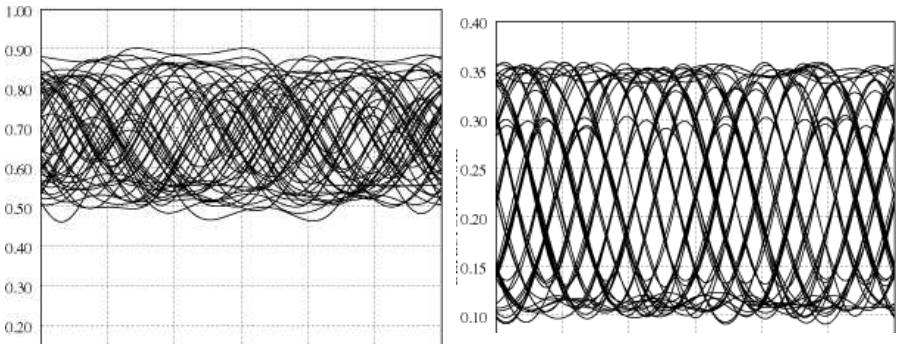

Fig. 22. A plot of the digital eye that is detected when focusing at a position 6 μm away from the paraxial focus (lefthand figure: far-field detection region is 80% of full aperture; righthand figure: 55% of the aperture is effectively used for detection. Note the loss in signal strength in the second picture.

ing digital eye pattern when the detection diameter in the far field is reduced from e.g. 80% to 55% at a focus setting with z=+6 μm from the paraxial focus P. In the case of a digital disc system, the quality of the detector signal is generally assessed by means of the so-called root mean square digital jitter. This quantity is obtained by binarising the analog signal read from a disc and comparing it with the desired binary wave form. The root mean square value of the shifts of the transients in the detected signal with respect to the corresponding positions in the originally recorded signal is called the digital jitter of the signal. The digital jitter Δ_d is expressed in the time domain or in the spatial domain and counted in units of the clock length T of the digital signal. In the righthand figure, the residual jitter amounts to 6.7%; this value is obtained in the case of a maximum decorrelation between neighbouring tracks. When the information between the tracks is fully correlated, the jitter goes down to 3% and in practice one finds values of 5.5 to 6% in optimum CD read-out conditions.

A further reduction in jitter is possible when the residual spherical aberration present over the central area is corrected in the objective by tailoring the surface profile of the first aspheric surface of the single element objective lens. In this case, a wave-front correction is applied to the central 55% region while the outer region is left unaltered and this leads to a further reduction of the jitter in the detected signal down to 5% for a CD-disc.

5.2 The effective optical transfer function

The changes at read-out by a reduction of the far-field detection area to e.g. 55 % of the total beam diameter can be expressed in terms of a modified MTF (modula-

tion transfer function) of the optical read-out system. A reduction of the detection aperture is equivalent to a more coherent or partially coherent detection of the information on the disc and this increases the response at the mid-spatial frequencies although, simultaneously, the cut-off frequency of the optics is reduced. This latter effect is not very important when reading a CD-disc at NA=0.60 and λ=650 nm because the cut-off frequency is far too high (1850 mm^{-1} instead of 1150 mm^{-1} for a standard CD-player).
In Fig.(23) we observe that at the optimum focus setting the resulting MTF (curves with triangles) is not much different from the standard MTF encountered in a classical CD-player with NA=0.45 and λ=785nm (the cut-off frequency would be found at an axial value of 1.15).
In the figure we also show the slight improvement that is possible when the residual

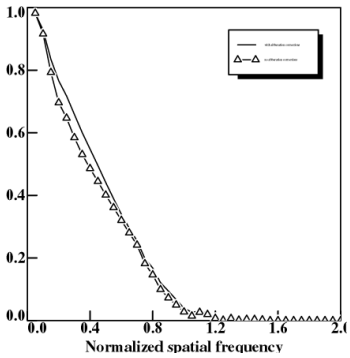

Fig. 23. MTF-curves of the read-out with reduced NA value at the detection side. The horizontal axis is expressed in units NA/λ. In the figure the optimum MTF obtained by selecting a relative aperture of 55% in the reflected beam (triangles) is further improved by introducing a correction for the residual spherical aberration over the central section of the objective (drawn line). For comparison, the optimum transfer function at full detection aperture has been sketched (dashed line).

spherical aberration present over the central area of the reading beam is corrected in the objective (drawn line). Some 5 to 10% improvement in MTF-value is possible at the mid CD-frequencies and this becomes visible also in the reduction of the bottom jitter. Moreover, the absence of residual spherical aberration over the effective detection area leads to a better symmetry of the read-out signal around the optimum focus setting.

5.3 The two-wavelength light path

As we remarked before, recordable or rewritable CD-discs require a read-out at the appropriate wavelength of λ=780-840 nm and the corresponding numerical aperture amounts to 0.45 . The partially coherent read-out at NA_D=55% that was possible at the red wavelength with a slight objective modification at the most, does not yield

a sufficiently large optical bandwidth at λ=780 nm. A solution is found by a further tailoring of a ring section of the surface profile of the first aspheric surface of the objective, thereby reducing the spherical aberration at CD read-out. The objective then can be considered to consist of three regions. The central paraxial region contributes

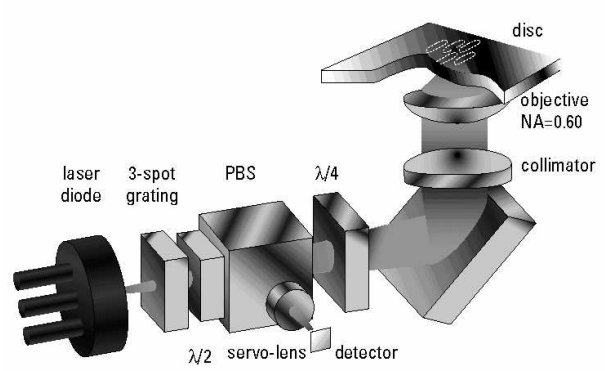

Fig. 24. The layout of an optical light path for DVD and CD with the possibility to inject an extra read/write beam of different colour via a second beam splitter (not shown in the figure), either before the beam splitter of in the parallel beam section after the collimator. The polarisation-sensitive light path with quarter-wave plate and polarising beam splitter (PBS) is used to maximise writing power.

to both DVD and CD read-out. A relatively narrow annular region corrects the CD read-out but is effectively lost for the DVD read-out. The outer region is only useful for the DVD read-out; this part of the read-out bundle becomes heavily aberrated at CD read-out and is then lost for detection [26]. The rather low light efficiency at CD read-out that was observed in Fig.(22) does not occur once a second independent source is used for CD read-out. The far-field of the second source can be adjusted to the maximum aperture ($0.45 \leq NA \leq 0.50$) needed for CD read-out.
Other compatibility solutions have been proposed based on a diffractive structure applied to the first aspheric surface of the objective and capable of correcting the spherical aberration in the first diffraction order [27]. More advanced solutions have recently been described in [28].

6 Efficient calculation scheme for the detector signal

In this section we will treat the problem of the efficient calculation of the detector signal in an optical disc player. The reliable and numerically efficient modelling of detected signals is a very welcome addition to the experimental verification of new ideas. With a reliable modelling tool, one obtains a quick estimate of tolerances and robustness of new options for recording schemes and read-out methods. In this section we present a numerical procedure that carries the mathematical analysis of

the read-out process as far as possible and thus gains precious time as compared to the purely numerical approach. Especially in the case of optics with general pupil shape and aberrations, the analytical approach can bring an important reduction in calculation time.

6.1 Optical configuration and the FFT-approach

The light path of an optical disc player is well represented by a scanning microscope of Type I according to the terminology given in [20]. In the schematic drawing of a light path according to Fig.(25), the detector could also be positioned at the location of the aperture of $O^{'}$ and in many modelling approaches, this is effectively done to minimize the numerical burden. But the way in which the high-frequency and the

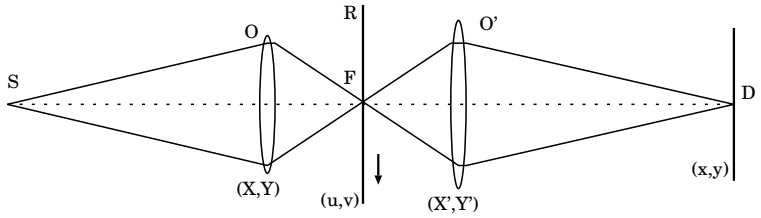

Fig. 25. Schematic drawing illustrating the progation of the light from the source S via the objective O towards the disc surface in the focal plane F of O. The disc presents a (spatially modulated) reflection function $R(u, v)$. After reflection the diffracted light is captured by the same objective (here denoted by $O^{'}$) and an image of the disc is produced at the location of the detector plane D. The pupil coordinates (X, Y) and $(X^{'}, Y^{'})$ are commonly normalized with respect to the half-diameter of the objective aperture. The disc coordinates (u, v) and the detector coordinates (x, y) are normalized with respect to the diffraction units λ/NA in both planes.

optical error signals are derived is more complicated and the diffraction step from the aperture of $O^{'}$ to the composite detectors in plane D is essential if a detailed knowledge of the intensity distribution on the detectors is needed. As an example we take the hologram beamsplitter of Fig.(11) where each beam propagating to the detector plane effectively originates from one half of the collecting aperture at $O^{'}$. Another example is the light path introducing astigmatism in Fig.(8). The intensity distribution on the detector in best focus is some 20 times larger in linear measure as compared to the standard diffraction image. It is important to know the distribution of the high-frequency signal in the various quadrants in both the in-focus situation and the defocused case.

The standard way to calculate a detector signal is the repeated application of a propagation step from an object or image plane to an aperture and vice versa, in optical terms by propagating from the near field to the far field and back. The degree of sophistication with which the propagation is carried out is not crucial in the sense that each propagation step can be reduced to a Fourier transform. Especially when

a scalar diffraction approach is sufficient, a single Fourier transform operation per step is adequate. For an efficient numerical implementation, the Fourier integral is replaced by a Fast Fourier Transform (FFT) that is faster with respect to execution time.
The assessment of the digital signal is carried out by means of the digital jitter Δ_d. As a rule of thumb, one can say that the rms jitter preferably should not exceed 10% of the clock length. At a value of 15%, prolonged over large signal portions, the digital signal fully breaks down and cannot be reconstructed without errors. The fact that some statistics are needed to calculate a reliable jitter value means that an extended track portion with a digitally modulated signal has to be scanned. For a reliable reconstruction of the optical signal, the sample points where the detector signal is calculated should be rather densely spaced, e.g. with a spatial increment of 0.05 to 0.1 of the diffraction unit λ/NA. For each sample point at the disc, the two FFT-steps from the disc to the aperture $O^{'}$ and to the detector plane D have to be executed, provided that the incident amplitude distribution at the focal point F is stored in memory. Several thousands of FFT-steps can be needed for obtaining one single value of Δ_d.

6.2 The analytic approach

The analysis given in this subsection uses the material presented in references [29] and [2]. In order to accommodate for more general signal sequences without loosing the advantage of dealing with periodic structures, we define a larger unit cell (length p) that contains several optical effects in the tangential direction and, if needed, several neighbouring tracks in the radial direction with a formal radial period of q (see Fig.(26)).
In the presence of the double periodic disc structure, the complex amplitude in the entrance aperture of the collecting objective $O^{'}$ (see Fig.(25)) is given by the sum of a certain number of diffracted waves according to

$$A'(X',Y') = \sum_{m,n} \rho_{m,\hat{n}} \ \exp\left\{2\pi i \left[\left(\frac{m}{p}\right) u + \left(\frac{\hat{n}}{q}\right) v\right]\right\} \\ f\ (X' - \frac{m}{p}, Y' - \frac{\hat{n}}{q})\ , \qquad (17)$$

where m is the tangential order number and the quantity $\hat{n}$ equals the radial order number $(n - ms/p)$ of a disc with a track-to-track shift s of the information. The argument of the exponential function is proportional to the mutual displacement (u, v) of the scanning spot and the disc structure in, respectively, the tangential and radial direction. The factor $\rho_{m,\hat{n}}$ is equal to the complex amplitude of the diffracted wave with the corresponding order number $(m, \hat{n})$. The complex function $f(X^{'}, Y^{'})$ stands for the lens function and carries the information about the lens transmission function (e.g. pupil shape, gaussian filling factor of the incoming beam) and the lens aberration.
Figure (27) shows the zeroth diffracted order and a general diffracted order (dotted

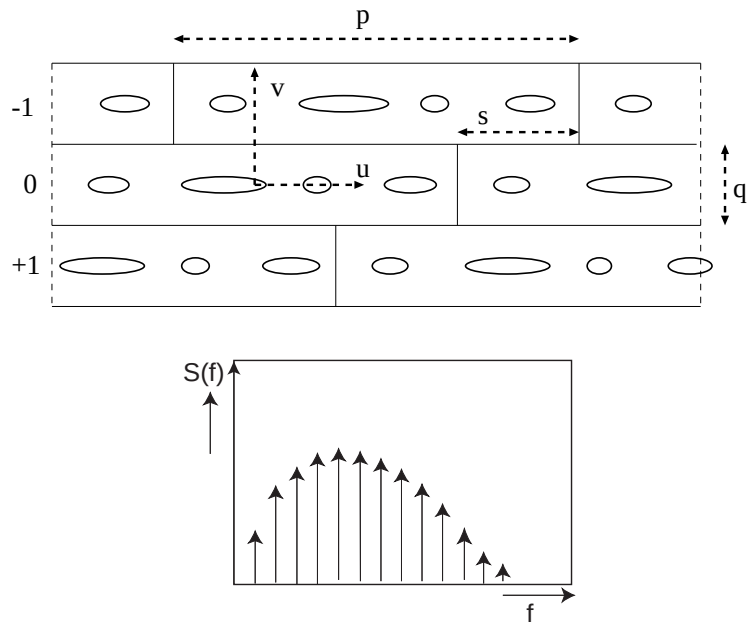

Fig. 26. The disc structure possesses a double periodicity. The fundamental frequency in the radial direction is given by $1/q$ with q in principle being equal to the pitch of the information track. The periodicity in the tangential direction is a 'synthetic' one. Within one period of length p several optical effects are present that obey the modulation rule of the optical structure with respect to the lengths of the optical effects and the land sections in between. The tangential information is repeated in the neighbouring tracks but a shift (distance s) has been provided to avoid directly correlated cross-talk. The tangential repetition with period p leads to a typical sampled power spectrum $S(f)$ of the digital information on the disc (lower graph); the distance between the frequency sample points along the normalised frequency axis is given by $\Delta f = \lambda/p(NA)$.

circles) as they are located in the pupil of the collecting objective $O^{'}$. The coordinates $(X^{'}, Y^{'})$ in the exit pupil are linearly related to the sines of the angles in the far-field of the disc structure if the collecting objective $O^{'}$ satisfies the sine condition of geometrical optics [30].
The imaging step from the exit pupil of $O^{'}$ to the detector plane is another Fourier transform from the coordinates $(X^{'}, Y^{'})$ to the detector plane coordinates (x, y). The lens function of $O^{'}$ is denoted by $g(X^{'}, Y^{'})$; in the special case of a reflective system, we put $g(X^{'}, Y^{'}) = f(-X, -Y)$. The complex amplitude $A^{''}(X^{'}, Y^{'})$ in the exit pupil of the collecting objective becomes the product of $A^{'}(X^{'}, Y^{'})$ and the lens function $g(X^{'}, Y^{'})$. If the disc is uniformly rotated, resulting in a local linear speed of s_0, the period $1/p$ is transformed into a temporal frequency $f_t = s_0/p$. We

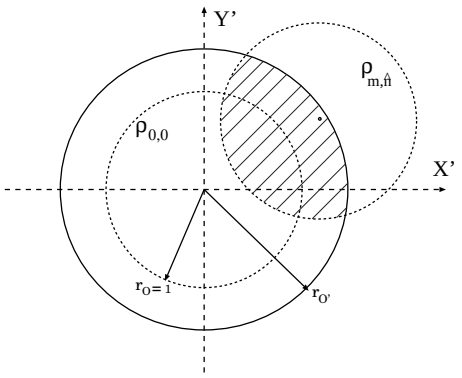

Fig. 27. Position of the diffracted orders of the disc structure with double periodicity in the exit pupil of the collecting objective $O^{'}$ with half diameter $r_{O'}$. The zeroth order with complex amplitude $\rho_{0,0}$ has been shown and a general diffraction order $(m, \hat{n})$. The hatched area indicates which part of the order $(m, \hat{n})$ is transmitted to the detector plane.

finally find for the complex amplitude in the detector plane

$$\begin{aligned} B(x,y;t) &= \int\int_{O'} A^{''}(X',Y') \, \exp\left\{2\pi i \left(X'x + Y'y\right)\right\} dX'dY' \\ &= \sum_{m,n} \rho_{m,\hat{n}} \, \exp\left\{2\pi i \left[m\left(f_t t - \frac{sv}{pq}\right) + \frac{nv}{q}\right]\right\} \\ &\quad \int\int_{O'} g\,(X',Y') f\,\left(X' - \frac{m}{p}\,, Y' - \frac{n - \frac{ms}{p}}{q}\right) \\ &\quad \exp\left\{2\pi i \left(X'x + Y'y\right)\right\} dX'dY'\,, \end{aligned} \tag{18}$$

where the integral is formally carried out over the full area of the exit pupil of $O^{'}$. In Eq.(18) we separate the integral over the exit pupil area $O^{'}$, that only depends on the imaging properties of the lenses, from the part that is scanning-dependent and carries information about the structure of the disc via the factors $\rho_{m,\hat{n}}$. The expression for $B(x,y;t)$ now reads

$$\begin{aligned} &B(x,y;t) \;= \\ &\sum_{m,n} \rho_{m,\hat{n}} \, \exp\left\{2\pi i \left[\left(m\left(f_t t - \frac{sv}{pq}\right) + \frac{nv}{q}\right)\right]\right\} F_{m,\hat{n}}(x,y)\,, \end{aligned} \tag{19}$$

where the factor $F_{m,\hat{n}}(x,y)$ equals the integral over X' and Y' in Eq.(18).
The detected quantity is the intensity, obtained by multiplying $B(x,y;t)$ with its complex conjugate and this yields

$$\begin{aligned} I(x,y;t) = |B(x,y;t)|^2 = \\ \sum_{m,n} \sum_{m',n'} \rho_{m,\hat{n}} \; {\rho^*}_{m',\hat{n}'} \; F_{m,\hat{n}}(x,y) \; {F^*}_{m',\hat{n}'}(x,y) \end{aligned}$$

$$\exp\left\{2\pi i\left[(m-m')\left(f_t t-\frac{sv}{pq}\right)+\frac{(n-n')v}{q}\right]\right\} \quad . \tag{20}$$

In an optical disc system we use special detector geometries, e.g. a quadrant detector. The detector signal is obtained by integrating the intensity over the detector area, taking into account a locally varying detection sensitivity function $S(x,y)$ and this yields a k^{th} detector signal according to

$$\begin{aligned} S_{D_k}(t) = & \sum_{m,n}\sum_{m',n'} \rho_{m,\hat{n}}\ {\rho^*}_{m',\hat{n}'} \\ & \exp\left\{2\pi i\left[(m-m')\left(f_t t-\frac{sv}{pq}\right)+\frac{(n-n')v}{q}\right]\right\} \\ & \int\int_{D_k} S(x,y)\ F_{m,\hat{n}}(x,y)\ {F^*}_{m',\hat{n}'}(x,y)dxdy \quad , \end{aligned} \tag{21}$$

where D_k denotes the area of the k^{th} detector. In the expression for the detector signal we have separated the quantities related to the disc structure from the integral that only contains information on the imaging process by the objective and collecting lens. This means that one single effort in evaluating the integral in Eq.(21) can serve for the many times repeated calculation of detector signals corresponding to different structures on the disc that fit into the basic period of length p. In this way we can gather statistical data on the digital jitter Δ_d with a seriously reduced numerical effort; in practice, a factor of the order of hundred is observed.

6.3 The harmonic components of the detector signal

With the double periodicity present on the disc, the detector signal can be written as a double Fourier series where the uniform scanning speed s_0 transforms the periodic components in the tangential direction into time harmonic components with a fundamental frequency $f_t = s_0/p$.
Using Eq.(21), we formally obtain the temporal harmonic components $A_{\mu,k}$ and $B_{\mu,k}$ from detector D_k according to

$$\begin{aligned} A_{0,k}(v) \quad = \quad & P_{1,k}(0,0) \\ & +2\sum_{\kappa=1}^{\kappa_{max}}\left\{P_{1,k}(0,\kappa)\cos\left(2\pi\frac{\kappa}{q}v\right)+P_{2,k}(0,\kappa)\sin\left(2\pi\frac{\kappa}{q}v\right)\right\} \\ A_{\mu,k}(v) \quad = \quad & 2\left\{P_{1,k}(\mu,0)\cos\left(2\pi\frac{s}{p}\frac{\kappa}{q}\mu\right)+P_{2,k}(\mu,0)\sin\left(2\pi\frac{s}{p}\frac{\kappa}{q}\mu\right)\right\} \\ & +2\sum_{\kappa=1}^{\kappa_{max}}\left\{\left[\{P_{1,k}(\mu,\kappa)+Q_{1,k}(\mu,\kappa)\}\cos\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right.\right. \\ & \qquad \left.+\{P_{2,k}(\mu,\kappa)+Q_{2,k}(\mu,\kappa)\}\sin\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right]\cos\left(2\pi\frac{\kappa}{q}v\right) \\ & \qquad +\left[\{P_{2,k}(\mu,\kappa)-Q_{2,k}(\mu,\kappa)\}\cos\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right. \end{aligned} \tag{22}$$

$$
\begin{aligned}
& -\{P_{1,k}(\mu,\kappa)-Q_{1,k}(\mu,\kappa)\}\sin\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\Big]\sin\left(2\pi\frac{\kappa}{q}v\right)\Big\} \\
B_{\mu,k}(v) \quad = \quad & 2\left\{-P_{2,k}(\mu,0)\cos\left(2\pi\frac{s}{p}\frac{\kappa}{q}\mu\right)+P_{1,k}(\mu,0)\sin\left(2\pi\frac{s}{p}\frac{\kappa}{q}\mu\right)\right\} \\
& +2\sum_{\kappa=1}^{\kappa_{max}}\left\{\left[\{-P_{2,k}(\mu,\kappa)-Q_{2,k}(\mu,\kappa)\}\cos\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right.\right. \\
& \left.+\{P_{1,k}(\mu,\kappa)+Q_{1,k}(\mu,\kappa)\}\sin\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right]\cos\left(2\pi\frac{\kappa}{q}v\right) \\
& +\left[\{P_{1,k}(\mu,\kappa)-Q_{1,k}(\mu,\kappa)\}\cos\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right. \\
& \left.\left.+\{P_{2,k}(\mu,\kappa)-Q_{2,k}(\mu,\kappa)\}\sin\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right]\sin\left(2\pi\frac{\kappa}{q}v\right)\right\} \ ,
\end{aligned}
$$

where the harmonic coefficients A and B, that depend on the off-track distance v of the scanning spot, generate the detector signal with the aid of the expression

$$
S_{D_k}(t,v)=A_{0,k}(v)+\sum_{\mu=1}^{\mu_{max}}\left\{A_{\mu,k}(v)\cos\left(\mu f_t t\right)+B_{\mu,k}(v)\sin\left(\mu f_t t\right)\right\} \ , \quad (23)
$$

where $\mu_{max} f_t$ is the maximum frequency transmitted by the optical read-out system.
The calculation of the detector signal $S_{D_k}(t,v)$ requires the evaluation of the coefficients $P_{j,k}(\mu,\kappa)$ and $Q_{j,k}(\mu,\kappa)$, $j=1,2$, that can be written as

$$
\begin{aligned}
P_{1,k}(\mu,\kappa) &= Re\left\{\sum_m\sum_n \rho_{\mu+m,\hat{\kappa}+\hat{n}}{\rho^*}_{m,\hat{n}} Z_k(\mu+m,\hat{\kappa}+\hat{n};m,\hat{n})\right\} \\
P_{2,k}(\mu,\kappa) &= Im\left\{\sum_m\sum_n \rho_{\mu+m,\hat{\kappa}+\hat{n}}{\rho^*}_{m,\hat{n}} Z_k(\mu+m,\hat{\kappa}+\hat{n};m,\hat{n})\right\} \\
Q_{1,k}(\mu,\kappa) &= Re\left\{\sum_m\sum_n \rho_{\mu+m,\hat{n}}{\rho^*}_{m,\hat{\kappa}+\hat{n}} Z_k(\mu+m,\hat{n};m,\hat{\kappa}+\hat{n})\right\} \\
Q_{2,k}(\mu,\kappa) &= Im\left\{\sum_m\sum_n \rho_{\mu+m,\hat{n}}{\rho^*}_{m,\hat{\kappa}+\hat{n}} Z_k(\mu+m,\hat{n};m,\hat{\kappa}+\hat{n})\right\} \quad (24)
\end{aligned}
$$

The factor $Z_k(m,\hat{n};m',\hat{n}')$ is the integral that was present in Eq.(21) and is written

$$
Z_k(m,\hat{n};m',\hat{n}')=\int\int_{D_k} S(x,y)\ F_{m,\hat{n}}(x,y)\ {F^*}_{m',\hat{n}'}(x,y)dxdy \ . \quad (25)
$$

The efficient calculation of the Z_k-functions is the subject of the remainder of this chapter. Especially the numerical evaluation of the basic function $F_{m,n}(x,y)$ can be time consuming and we will show how a further analysis of this function can improve the efficiency of the final evaluation.

6.4 The representation of the function $F_{m,n}(x, y)$

The function $F_{m,n}(x, y)$ is the Fourier Transform of the product of the displaced objective lens function $f(X^{'} - m/p, Y^{'} - n/q)$ and the collecting lens function $g(X^{'}, Y^{'})$ where the integration area is defined by the half diameter $r_{O'}$ of the collecting objective ($r_{O'}$ is commonly normalised to unity). The standard way to calculate the Fourier transform is using a numerical two-dimensional fast Fourier Transform (FFT). In this paragraph we will show that the use of orthogonal functions in the diffracting exit pupil and in the image plane where the detectors are situated can lead to an explicit expression for the detector intensity. This expression is then readily integrated over the detector area defined by D_k.
The function to be Fourier transformed generally is a complex function that is nonzero on a typical domain within the exit pupil of $O^{'}$ like the hatched area in Fig.(27). Orthogonal functions with uniform weighting on the circular exit pupil area of $O^{'}$ (pupil radius normalised to unity) are the well-known Zernike polynomials [31]. These polynomials are normally used to represent the argument of the complex pupil function, in particular the aberration or the focus defect of an imaging optical system. Here we propose to extend the use of the Zernike polynomials and to represent the complete complex function $A(\rho, \theta) \exp[i\Phi(\rho, \theta)]$ in the exit pupil by a set of Zernike polynomials yielding complex expansion coefficients a_{kl} according to

$$A(\rho, \theta) \exp[i\Phi(\rho, \theta)] = \sum_{k,l} \alpha_{kl}\ R_k^l(\rho) \cos l\theta\ , \tag{26}$$

where (ρ, θ) are the polar coordinates in the exit pupil, $A(\rho, \theta)$ the amplitude and $\Phi(\rho, \theta)$ the phase of the pupil function. $R_k^l(\rho)$ is the radial part of the Zernike polynomial with radial degree k and azimuthal degree l ($k \geq l \geq 0$, $k - l$ even). We have limited ourselves to the cosine-polynomials, thus restricting the functions to those that are symmetric with respect to $\theta = 0$. The expansion of a general function would require an extra set of coefficients b_{kl} associated with the sine-polynomials.
The expansion in Eq.(26) of $A \exp(i\Phi)$ can be obtained, for instance, by a least squares fit with a finite series of polynomials, a procedure which is common practice for expanding Φ itself. It can be shown that, with the transition from cartesian coordinates (x, y) to normalised polar coordinates (r, ϕ), the complex amplitude U in the detector plane is given by

$$U(r, \phi) = 2 \sum_{k,l}\ a_{kl}\ i^l\ V_{kl} \cos l\phi \quad , \tag{27}$$

where

$$V_{kl} = \int_0^1 \rho \exp(i f_z \rho^2) R_k^l(\rho) J_l(2\pi\rho r) d\rho \tag{28}$$

for integers $k, l \geq 0$ with $k - l \geq 0$ and even (J_l denotes the Bessel function of the first kind of order l). A quadratic phase factor $\exp(i f_z \rho^2)$ has been added that is needed in the case of a defocusing of the image plane with respect to the optimum

(paraxial) focus.
The Bessel series representation of V_{kl} is given in [32], [33] and reads

$$V_{kl} = \exp(if_z) \sum_{s=1}^{\infty} (-2if_z)^{s-1} \sum_{j=0}^{p_z} v_{sj} \frac{J_{l+s+2j}(t)}{st^s} \tag{29}$$

with v_{sj} given by

$$v_{sj} = (-1)^{p_z}(l+s+2j) \binom{l+j+s-1}{s-1} \binom{j+s-1}{s-1} \binom{s-1}{p_z-j} \Big/ \binom{q_z+s+j}{s} , \tag{30}$$

for $s = 1, 2, \ldots;\ j = 0, \ldots, p_z$. In Eq.(29) we have set

$$t = 2\pi r \ , \qquad p_z = \frac{k-l}{2} \ , \qquad q_z = \frac{k+l}{2} \ . \tag{31}$$

For the number S of terms to be included in the infinite series over s we have the following rule. It can be shown that, when S=25, the absolute truncation error is of the order 10^{-6} for all f_z, t, k, l specified by

$$|f_z| \leq 2\pi, \quad t \leq 20, \quad 0 \leq p_z \leq q_z \leq 6 \ . \tag{32}$$

In the absence of defocusing ($f_z = 0$), the expansion of the complex pupil function in terms of Zernike polynomials leads to the relationship

$$U(r,\phi) = 2 \sum_{k,l} a_{kl}\ i^l\ (-1)^{\frac{k-l}{2}}\ \frac{J_{k+1}(2\pi r)}{2\pi r}\ \cos l\phi \ . \tag{33}$$

This analytic result for the in-focus amplitude can be obtained using formula (39), p. 772 in [31]

$$\int_0^1 \rho R_k^l(\rho) J_l(2\pi\rho r) d\rho = (-1)^{\frac{k-l}{2}}\ \ \frac{J_{k+1}(2\pi r)}{2\pi r}. \tag{34}$$

but it also follows as the limiting case for $f_z \to 0$ of Eqs.(28)-(29).

6.5 Orthogonality in pupil and image plane

Having seen the one-to-one relationship between the orthogonal Zernike polynomials in the pupil and the corresponding Bessel functions of the first kind in the image plane (in-focus situation), we are left with the evaluation of the inner products of the Bessel functions to check their orthogonality and establish a normalisation factor. We define the inner product of two functions from the series in Eq.(33) as

$$\begin{aligned} I_{k,l;k',l'} &= C \int_0^{2\pi} \int_0^{\infty} U_{kl} U_{k'l'}\, r dr d\phi \\ &= 4C\ i^{k+k'} \int_0^{2\pi} \int_0^{\infty} \frac{J_{k+1}(2\pi r) J_{k'+1}(2\pi r)}{4\pi^2 r^2} \cos l\phi \cos l'\phi\ r dr d\phi \\ &= \frac{C}{\pi}\ i^{k+k'} \int_0^{\infty} \frac{J_{k+1}(2\pi r) J_{k'+1}(2\pi r)}{r}\ dr\ \ \delta(l, l') \ , \end{aligned} \tag{35}$$

with $\delta(l,l^{'})$ the Kronecker symbol ($=2$ if $l=l^{'}$=0, $=1$ if $l=l^{'}\neq$0 and equals 0 if $l\neq l^{'}$). The common conditions for the indices of Zernike polynomials apply too (both $k-l$ and $k^{'}-l^{'}$ are even).
For the calculation of the definite integral over r we use the special result for the integral of the product of two Bessel functions in paragraph 11.4.6 of [34]

$$\int_0^\infty t^{-1}J_{\nu+2n+1}(t)J_{\nu+2m+1}(t)dt = 0 \qquad (m\neq n)$$
$$= \frac{1}{2(2n+\nu+1)} \quad (m=n)\ , \tag{36}$$

with the restriction $\nu+n+m>-1$.
Applied to Eq.(35) with $\nu+2n=k$, $\nu+2m=k^{'}$ and $t=2\pi r$, we find the result

$$I_{k,l;k^{'},l^{'}} = \frac{C}{\pi} \qquad (k=k^{'}=l=l^{'}=0)$$
$$= \frac{(-1)^k}{2(k+1)} \qquad (k=k^{'}\neq 0;\ l=l^{'}\neq 0) \tag{37}$$

For normalisation purposes, we multiply the basic functions with $\sqrt{\pi}$ so that the value of $I_{00;00}$ equals unity.

The evaluation of the function $Z_k(m,\hat{n};m',\hat{n}')$

The quantity $Z_k(m,\hat{n};m',\hat{n}')$ was the key quantity to be calculated for the evaluation of the harmonic signal components (see Eq.(25)).
Using the orthogonal expansion for the functions $F_{m,n}(x,y)$ in the exit pupil of the objective $O^{'}$ according to Eq.(33) we obtain

$$Z_j(m,\hat{n};m',\hat{n}') = \sum_{kl}\sum_{k'l'} a_{kl}a^{'*}_{k'l'}\ i^{l-l^{'}}\ (-1)^{\frac{k+k^{'}-l-l^{'}}{2}}$$
$$\int\int_{D_j} S_r(r,\phi) \quad \frac{J_{k+1}(2\pi r)J_{k^{'}+1}(2\pi r)}{4\pi^2r^2}\cos l\phi\cos l^{'}\phi\ rdrd\phi\ . \tag{38}$$

where the asterisk denotes the complex conjugate function and the detector sensitivity function S_r has been written in polar coordinates.
In the general situation with an arbitrary detector sensitivity function, the integrals above have to be evaluated numerically. In the whole analytic process described until now, the calculation of the coefficients a_{kl} of the Zernike expansion of the complex function in the exit pupil of $O^{'}$ is the crucial step with respect to speed and convergence and this aspect is actively researched at this moment.

7 Conclusion

In this chapter, we have described some important aspects of the optical disc systems that are actually widely used, the CD- and the DVD-system. The subject of optical storage being vast, we have treated some restricted topics that are of interest

for the understanding of the CD- and DVD-system themselves and the exchange of media between these two standardized platforms of optical recording. A detailed analysis has been given of the signal detection in an optical disc system. This is of particular interest when simulations are used as a predictive tool, in parallel to experimental research. The efficiency of the calculations is greatly enhanced when the data belonging to the information structure on the disc are separated from the imaging process in the optical recording unit; a procedure to achieve this separation has been described in this chapter and an important reduction in calculation time is obtained. We have also indicated that, by using Zernike polynomials to represent complex amplitude functions, the analytic calculation process can be pursued further and the speed in evaluating the detector signal is again improved.
Because of space limitations, we have not discussed recent developments in optical storage like the blue laser DVR-system and still other research activities like near-field optical recording that aim at super high-density discs. With respect to these new developments, the analyses given in this chapter have to be adapted to the vectorial diffraction aspects of high-aperture image formation.

References

[1] G. Bouwhuis and J. Braat: "Recording and reading of information on optical disks" in *Applied Optics and Optical Engineering*, Vol. 9, ed. by R.R. Shannon and J.C. Wyant (Academic, New York 1983)

[2] G. Bouwhuis, J. Braat, A. Huijser, J. Pasman, G. van Rosmalen and K. Schouhamer Immink: *Principles of optical disc systems* (Adam Hilger, Bristol 1984)

[3] J. Isailovic: *Videodisc and optical memory systems* (Prentice Hall, Englewood Cliffs 1985)

[4] A. Marchant: *Optical recording; a technical overview* (Addison Wesley 1990)

[5] *Selected Papers on Optical Storage*, in *SPIE Milestone Series*, Vol. MS **49**, ed. G. Sincerbox and J. Zavislan (SPIE Optical Engineering Press, Bellingham 1992)

[6] K. Schwartz: *The physics of optical recording* (Springer, Berlin 1993)

[7] M. Mansuripur: *The physical principles of magneto optical data recording* (Cambridge University Press, Cambridge 1995)

[8] E.W. Williams: *Textbooks in electrical and electronic engineering Vol 2, The CD-ROM and optical recording systems* (Oxford Univ. Press 1996)

[9] S. Stan: *The CD-ROM drive, a brief system description* (Kluwer Academic Publishers, Dordrecht 1998)

[10] M. Minsky: *Microscopy Apparatus* U.S. Patent 3,013,467 (1961)

[11] Ph. Rice et al.: J. Soc. Motion Pict. Telev. Eng. **79**, 997 (1970)

[12] K. Compaan, P. Kramer: Philips Tech. Rev. **33**, 178 (1973)

[13] G. Broussaud, E. Spitz, C. Tinet and F. LeCarvenec: IEEE Trans. Broadcast Telev. Receivers **BTR-20**, 332 (1974)

[14] G.W. Hrbek: J. Soc. Motion Pict. Telev. Eng. **83**, 580 (1974)

[15] J. Braat: Appl. Opt. **22**, 2196 (1983)

[16] J. Haisma, E. Hugues and C. Babolat: Opt. Lett. **4(2)**, 70 (1979)
[17] M. Terao, T. Nishida, Y. Miyauchi, S. Horigome, T. Kaku and N. Ohta: Proc. Soc. Photo-Opt. Instrum. Eng. **695**, 105 (1986)
[18] T. Ohta, K. Inoue et al.: Jpn. J. Appl. Phys. **28** (Suppl. 28-3), 123 (1989)
[19] W.T. Welford: Journal of Microscopy **96**, 105 (1972)
[20] T. Wilson, C.J.R. Sheppard: *Theory and practice of the scanning optical microscope* (Cambridge Univ. Press, Cambridge 1984)
[21] C. Bricot, J. Lehureau, C. Puech and F. Le Carvennec: IEEE Trans. Consum. Electron. CE-**22**, 304 (1976)
[22] M. Mansuripur and C. Pons: Proc. Soc. Photo-Opt. Instrum. Eng. **899**, 56 (1988)
[23] G. Bouwhuis and P. Burgstede: Philips Tech. Rev. **33**, 186 (1973)
[24] J. Braat, G. Bouwhuis: Appl. Opt. **17**, 2013 (1978)
[25] J. Heemskerk, K. Schouhamer Immink: Philips Tech. Rev. **40**, 157 (1982)
[26] N. Arai,H.Yamazaki, S. Saito: U.S. Patent 6,243,349 B1
[27] Y. Komma, K. Kasazumi, S. Nishino and S. Mizumo: Proc. Soc. Photo-Opt. Instrum. Eng. **2338** 282-288 (1994)
[28] B.H.W. Hendriks, J.E. de Vries and H.P. Urbach: Appl. Opt. **40**, 6548 (2001)
[29] H.H. Hopkins: J. Opt. Soc. Am **69**, 4 (1979)
[30] W.T. Welford: *The aberrations of optical systems* (Adam Hilger, Bristol 1986)
[31] M. Born, E. Wolf: *Principles of Optics* (Pergamon Press, Oxford 1970)
[32] A.J.E.M. Janssen: J. Opt. Soc. Am. A **19**, May issue (2002)
[33] J. Braat, P. Dirksen, A.J.E.M. Janssen: J. Opt. Soc. Am. A **19**, May issue (2002)
[34] M. Abramowitz, I. Stegun: *Handbook of Mathematical Functions* (9th printing, Dover, New York 1970)