3GE Collection on Computer Science:

3D Optical Storage Technology

BIBLIOTEX
Digital Library

3GE COLLECTION ON COMPUTER SCIENCE:

3D OPTICAL STORAGE TECHNOLOGY

BIBLIOTEX
Digital Library

www.bibliotex.com

3GE COLLECTION ON COMPUTER SCIENCE:
3D OPTICAL STORAGE TECHNOLOGY

www.bibliotex.com

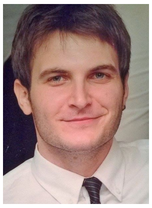
email: info@bibliotex.com

e-book Edition 2022

ISBN: 978-1-98467-702-0 (e-book)

In Collaboration with 3G E-Learning LLC. Originally Published in printed book format by 3G E-Learning LLC with ISBN 978-1-98465-995-8

EDITORIAL BOARD



Aleksandar Mratinković is born on May 5, 1988. in Arandjelovac, Serbia. He has graduated on Economic high school (2007), The College of Tourism in Belgrade (2013), and also has a master degree of Psychology (Faculty of Philosophy, University of Novi Sad). He has been engaged in different fields of psychology (Developmental Psychology, Clinical Psychology, Educational Psychology and Industrial Psychology) and has published several scientific works.

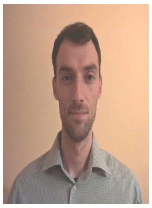
Dan Piestun (PhD) is currently a startup entrepreneur in Israel working on the interface of Agriculture and Biomedical Sciences and was formerly president-CEO of the National Institute of Agricultural Research (INIA) in Uruguay. Dan is a widely published scientist who has received many honours during his career including being a two-time recipient of the Amit Golda Meir Prize from the Hebrew University of Jerusalem, his areas of expertise includes stem cell molecular biology, plant and animal genetics and bioinformatics. Dan's passion for applied science and technological solutions did not stop him from pursuing a deep connection to the farmer, his family and nature. Among some of his interest and practices counts enjoying working as a beekeeper and onboard fishing.

Felecia Killings is the Founder and CEO of LiyahAmore Publishing, a publishing company committed to providing technical and educational services and products to Christian Authors. She operates as the Senior Editor and Writer, the Senior Writing Coach, the Content Marketing Specialist, Editor-in-Chief to the company's quarterly magazine, the Executive and Host of an international virtual network, and the Executive Director of the company's online school for Authors. She is a former high-school English instructor and professional development professor. She possesses a Master of Arts degree in Education and a Bachelor's degree in English and African American studies.

Dr. Sandra El Hajj, Ph.D. in Health Sciences from Nova Southeastern University, Florida, USA is a health professional specialized in Preventive and Global Health. With her 12 years of education obtained from one of the most prominent universities in Beirut, in addition to two leading universities in the State of Florida (USA), Dr. Sandra made sure to incorporate interdisciplinary and multicultural approaches in her work. Her long years of studies helped her create her own miniature world of knowledge linking together the healthcare field with Medical Research, Statistics, Food Technology, Environmental & Occupational Health, Preventive Health and most noteworthy her precious last degree of Global Health. Till today, she is the first and only doctor specialized in Global Health in the Middle East area.

Hazem Shawky Fouda has a PhD. in Agriculture Sciences, obtained his PhD. From the Faculty of Agriculture, Alexandria University in 2008, He is working in Cotton Arbitration & Testing General Organization (CATGO).

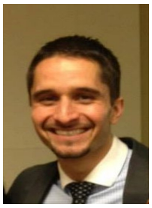
Fozia Parveen has a Dphil in Sustainable Water Engineering from the University of Oxford. Prior to this she has received MS in Environmental Sciences from National University of Science and Technology (NUST), Islamabad Pakistan and BS in Environmental Sciences from Fatima Jinnah Women University (FJWU), Rawalpindi.



Igor Krunic 2003-2007 in the School of Economics. After graduating in 2007, he went on to study at The College of Tourism, at the University of Belgrade where he got his bachelor degree in 2010. He was active as a third-year student representative in the student parliament.Then he went on the Faculty of science, at the University of Novi Sad where he successfully defended his master's thesis in 2013. The crown of his study was the work titled Opportunities for development of cultural tourism in Cacak". Later on, he became part of a multinational company where he got promoted to a deputy director of logistic. Nowadays he is a consultant and writer of academic subjects in the field of tourism.

Dr. Jovan Pehcevski obtained his PhD in Computer Science from RMIT University in Melbourne, Australia in 2007. His research interests include big data, business intelligence and predictive analytics, data and information science, information retrieval, XML, web services and service-oriented architectures, and relational and NoSQL database systems. He has published over 30 journal and conference papers and he also serves as a journal and conference reviewer. He is currently working as a Dean and Associate Professor at European University in Skopje, Macedonia.

Dr. Tanjina Nur finished her PhD in Civil and Environmental Engineering in 2014 from University of Technology Sydney (UTS). Now she is working as Post-Doctoral Researcher in the Centre for Technology in Water and Wastewater (CTWW) and published about eight International journal papers with 80 citations. Her research interest is wastewater treatment technology using adsorption process.



Stephen obtained his PhD from the University of North Carolina at Charlotte in 2013 where his graduate research focused on cancer immunology and the tumor microenvironment. He received postdoctoral training in regenerative and translational medicine, specifically gastrointestinal tissue engineering, at the Wake Forest Institute of Regenerative Medicine. Currently, Stephen is an instructor for anatomy and physiology and biology at Forsyth Technical Community College.

Michelle holds a Masters of Business Administration from the University of Phoenix, with a concentration in Human Resources Management. She is a professional author and has had numerous articles published in the Henry County Times and has written and revised several employee handbooks for various YMCA organizations throughout the United States.

Table of Contents

Preface

The best medium for data storage has been a main concern over the past decades. The advent of the information age is accompanied by the rapid growth of digital information, which is in urgent need for the development of huge storage space and security media. Optical storage technology with the huge storage capacity and low cost has become a new choice for information storage. At present, the commercial optical storage devices mainly include Blu-ray discs, digital versatile discs, and compact discs. The main digital circulation has been the optical data storage which has been being improved to accommodate new changes in technology and applications. Optical data storage is an alternative to magnetic disk data storage. Currently data access times are extremely slow for magnetic disks when compared to the speed of execution of CPUs so that any improvement in data access speeds will greatly increase the capabilities of computers, especially with large data and multimedia files. Optical memory is a technology that uses a three dimensional medium to store data and it can access such data a page at a time instead of sequentially, which leads to increases in storage density and access speed. Optical data storage systems are very close to becoming economically feasible. Photo-refractive crystals and photopolymers have been used successfully in experimental optical data storage systems. Such systems exploit the optical properties of these photosensitive materials along with the behavior of laser light when it is used to record an image of an object. 3-D optical data storage technology is one of the modern methods of storing large volumes of data.

This book discusses in details the fundamentals of 3D optical data storage. This includes the features of the 3D optical data storage and the major components that make up the devices. Nonresonant Multiphoton, Sequential multiphoton absorption, microholography and data recording are some of the writing methods used in the 3D optical data storage. The major challenges that are facing these devices as discussed in the book are; media sensitivity, Thermodynamic stability and destructive reading. Two volumetric optical storage approaches, holographic storage and localized-bit storage, were discussed. These approaches provide better capacity, input and output data rates, latency, system volume and low cost in comaprison with the conventional approaches. Storage and retrieval of long data in a relatively smaller space is a challenging task for communication engineer. Now a day’s CD’s, DVD’s, pen derives and hard disk are usually used for this purpose which are not capable holding large amount of data and also retrieval of data takes relatively last time. This book is intended to review the storage of data in 3D optical medium which will hold the large amount of data and will make retrieval easier. The book will help you to study about 3D optical data storage, what is data recording, what is its process, and comparisons with holographic data storage.

CHAPTER 1

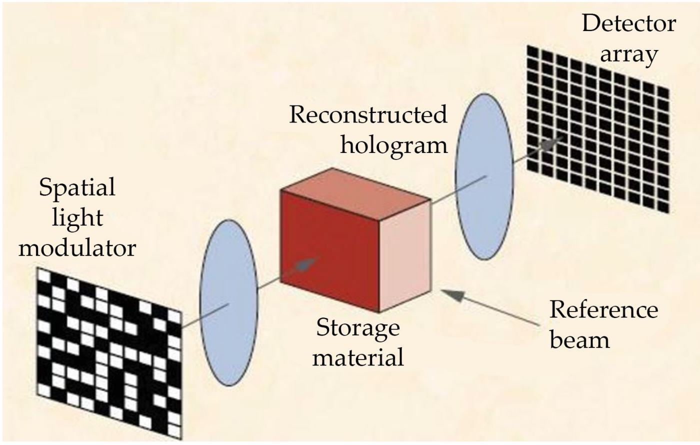
3D OPTICAL DATA STORAGE

INTRODUCTION

3D optical data storage is the term given to any form of optical data storage in which information can be recorded and/or read with three dimensional resolution (as opposed to the two dimensional resolution afforded, for example, by CD).

This innovation has the potential to provide petabyte-level mass storage on DVD-sized disks. Data recording and readback are achieved by focusing lasers within the medium. However, because of the volumetric nature of the data structure, the laser light must travel through other data points before it reaches the point where reading or recording is desired. Therefore, some kind of nonlinearity is required to ensure that these other data points do not interfere with the addressing of the desired point.

No commercial product based on 3D optical data storage has yet arrived on the mass market, although several companies are actively developing the technology and claim that it may become available soon.



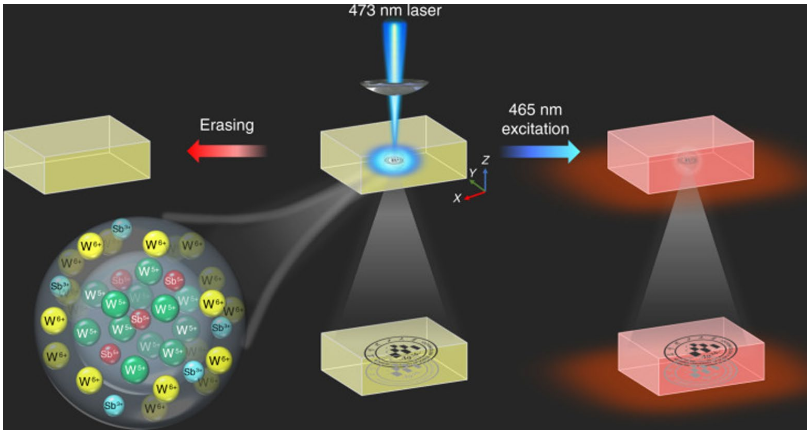
1.1 BASICS OF 3D OPTICAL DATA STORAGE

3D optical data storage is the term given to any form of optical data storage in which information can be recorded and/or read with three dimensional-resolution (as opposed to the two dimensional resolution afforded, for example, by CD).

This innovation has the potential to provide petabyte-level mass storage on DVD-sized disks. Data recording and read-back are achieved by focusing lasers within the medium. However, because of the volumetric nature of the data structure, the laser light must travel through other data points before it reaches the point where reading or recording is desired.

Therefore, some kind of nonlinearity is required to ensure that these other data points do not interfere with the addressing of the desired point. No commercial product based on 3D optical data storage has yet arrived on the mass market, although several companies are actively developing the technology and claim that it may become available soon.

The origins of the field date back to the 1950s, when Yehuda Hirshberg developed the photochromicspiropyrans and suggested their use in data storage. In the 1970s, Valeri Barachevskii demonstrated that this photochromism could be produced by two-photon excitation, and finally at the end of the 1980s Peter T. Rentzepis showed that this could lead to three-dimensional data storage. This proof-of-concept system stimulated a great deal of research and development, and in the following decades many academic and commercial groups have worked on 3D optical data storage products and technologies. Most of the developed systems are based to some extent on the original ideas of Rentzepis.



A wide range of physical phenomena for data reading and recording have been investigated, large numbers of chemical systems for the medium have been developed and evaluated, and extensive work has been carried out in solving the problems associated with the optical systems required for the reading and recording of data. Currently, several groups remain working on solutions with various levels of development and interest in commercialization.

1.1.1 Features of Three-Dimensional Optical Data Storage

Different from 2Ddata storage where data is stored on an aluminum film built in a disk and read through focusing of laser beams on its surface, 3D optical data storage uses nonlinear methods of reading and recording data. In 3D, data storage and reading are facilitated through the focusing of a laser on the storage medium where in this case, the laser passes through other points as it moves in a nonlinear way before it reaches the point where data recording or reading is wanted. This is an improvement of the 2D optical data storage where the laser light bypasses every layer of data stored in the storage medium before reaching the desired point leading to the recording of only 10 layers of data which is a limitation of the current technology. 3D optical data storage uses methods in which the light interacts with only the addressed volumetric pixel. In this method, laser light also passed through other data points but due to its nonlinearity, it has no effect on them. 3D optical data storage has the ability to provide mass storage of up to 1024 TB on a single DVD disk which beyond the current 2D optical storage.

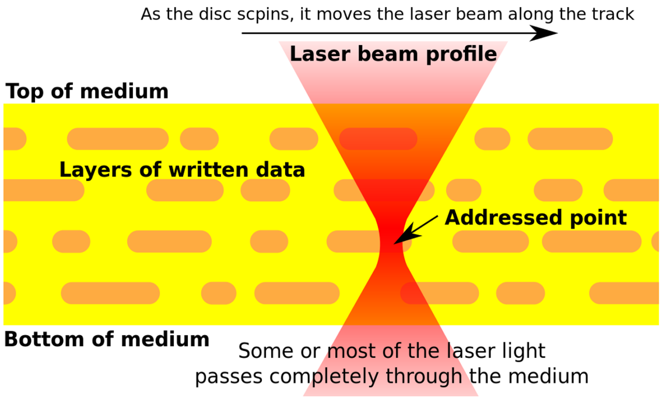
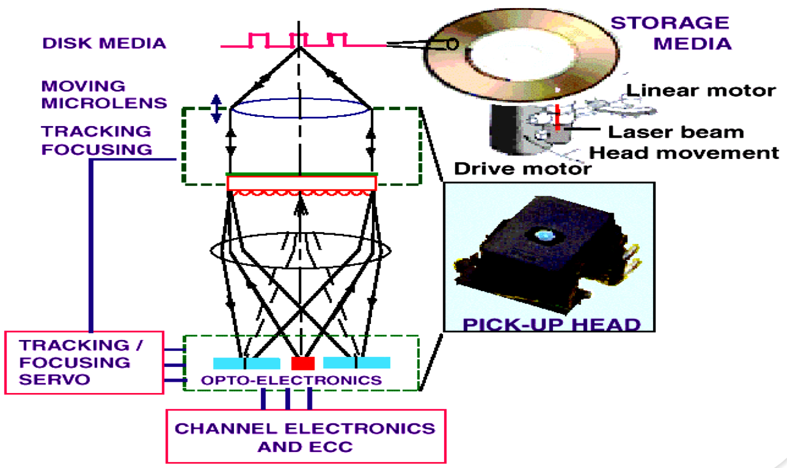


Figure 1: A schematic diagram showing a cross-section of an optical 3D storage.

1.1.2 Optical Recording Technology

Optical storage systems consist of a drive unit and a storage medium in a rotating disk form. In general the disks are pre-formatted using grooves and lands (tracks) to enable the positioning of an optical pick-up and recording head to access the information on the disk. Under the influence of a focused laser beam emanating from the optical head, information is recorded on the media as a change in the material characteristics. The disk media and the pick-up head are rotated and positioned through drive motors controlling the position of the head with respect to data tracks on the disk. Additional peripheral electronics are used for control and data acquisition and encoding/decoding.



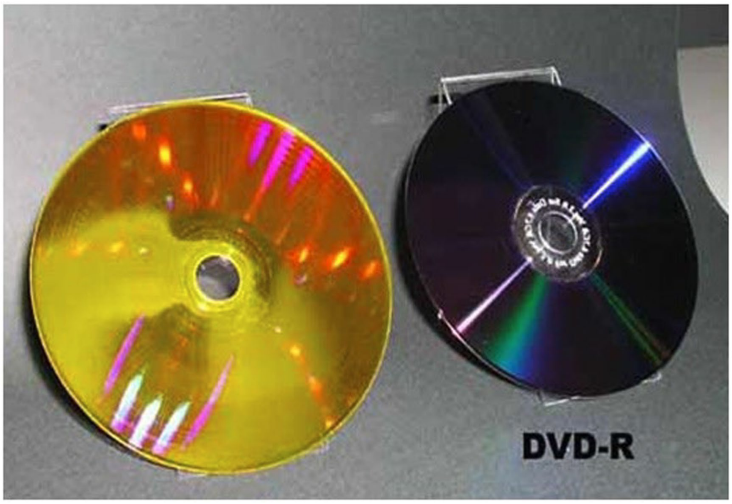
As an example, a prototypical 3D optical data storage system may use a disk that looks much like a transparent DVD. The disc contains many layers of information, each at a different depth in the media and each consisting of a DVD-like spiral track. In order to record information on the disc a laser is brought to a focus at a particular depth in the media that corresponds to a particular information layer. When the laser is turned on it causes a photochemical change in the media. As the disc spins and the read/write head moves along a radius, the layer is written just as a DVD-R is written. The depth of the focus may then be changed and another entirely different layer of information written. The distance between layers may be 5 to 100 micrometers, allowing >100 layers of information to be stored on a single disc.

In order to read the data back (in this example), a similar procedure is used except this time instead of causing a photochemical change in the media the laser causes fluorescence. This is achieved e.g. by using a lower laser power or a different laser wavelength. The intensity or wavelength of the fluorescence is different depending on whether the media has been written at that point, and so by measuring the emitted light the data is read.

The size of individual chromophore molecules or photoactive color centers is much smaller than the size of the laser focus (which is determined by the diffraction limit). The light therefore addresses a large number (possibly even 109) of molecules at any one time, so the medium acts as a homogeneous mass rather than a matrix structured by the positions of chromophores

1.1.3 Comparison with Holographic Data Storage

3D optical data storage is related to (and competes with) holographic data storage. Traditional examples of holographic storage do not address in the third dimension, and are therefore not strictly "3D", but more recently 3D holographic storage has been realized by the use of microholograms. Layer-selection multilayer technology (where a multilayer disc has layers that can be individually activated e.g. electrically) is also closely related.



Holographic data storage is a potential replacement technology in the area of high-capacity data storage currently dominated by magnetic and conventional optical data storage. Magnetic and optical data storage devices rely on individual bits being stored as distinct magnetic or optical changes on the surface of the recording medium. Holographic data storage overcomes this limitation by recording information throughout the volume of the medium and is capable of recording multiple images in the same area utilizing light at different angles.

Additionally, whereas magnetic and optical data storage records information a bit at a time in a linear fashion, holographic storage is capable of recording and reading millions of bits in parallel, enabling data transfer rates greater than those attained by traditional optical storage.

The stored data is read through the reproduction of the same reference beam used to create the hologram. The reference beam's light is focused on the photosensitive material, illuminating the appropriate interference pattern, the light diffracts on the interference pattern, and projects the pattern onto a detector. The detector is capable of reading the data in parallel, over one million bits at once, resulting in the fast data transfer rate. Files on the holographic drive can be accessed in less than 200 milliseconds.

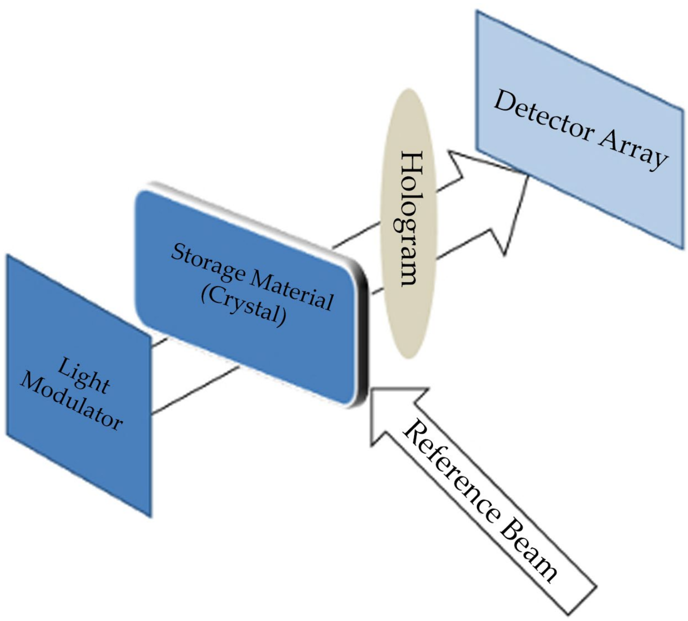
1.1.4 Types of 3D storage

There are two (2) major types of 3D storage:

- The simple storage of data throughout the volume of the disc; and
- Holographic storage.

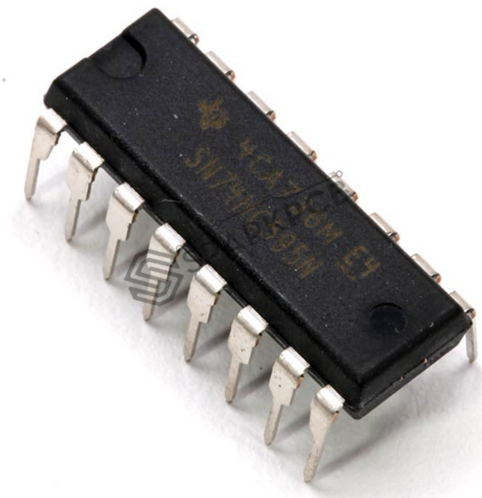
3–D Storage of Localized Bits

The most straightforward version of volumetric optical storage is the intuitive extension from surface optical storage: localized bits stored not only on the surface but throughout the volume. In relation to the abstract black–box, the modulator is usually the laser itself, and the selective addressing of data–sets is done by focusing the laser beam. Researchers have been exploring several variants of bit–localized storage, which can be roughly grouped into proposals which read one bit at a time, and those which can read multiple bits in parallel. With the former, the laser focuses to single voxels, and reads data out to a single photodetector (or differentially using a few detectors); with the latter, the laser selects a small set of contiguous voxels, and then reads data out to a photodetector array (a CCD or CMOS camera).



Bit–Serial Storage

Extending the CD concept to multiple layers can be done without changing much of the readout hardware. The focus servo becomes responsible for changing between layers of different depth, in addition to its primary task of locking onto a layer once one is chosen. Crosstalk from other layers (and the possibility of confusion for the focus servo) is minimized by separating layers by a fairly large distance as shown in Figure 2 (this spacing is ~ 55 microns for DVD–ROM). This ensures that when the converging (diverging) beam passes through the nearest neighboring layers, the large spot size covers enough data bits that the loss in transmission due to reflecting pits (and the out-of-focus crosstalk signal in reflection) remains roughly constant. If the layers are moved closer together, more crosstalk reaches the detector, and the smaller pool of illuminated data bits means that the statistical variation of random ON and OFF bits can become a significant noise source.



As more layers are added, then the reflectance and transmission of each layer needs to be adjusted so that the signals from each layer are equally detectable. (In the DVD–ROM, this is done by using a gold coating on the top layer and aluminum on the bottom.) The signal-to-noise ratio is reduced not only by the lower average signal level, but also by the scattering of the reflected beam as it passes through higher layers on its way back to the detector. As the number of layers increases and the bottom layers move relatively deep into the substrate, a tradeoff emerges between high numerical aperture (needed for tight focusing more than depth of focus since the disk is layered already) and the working distance between the lens and disk surface. Spherical aberration, usually corrected for in the optics, now varies from layer to layer. Without adaptive correction for each layer, this rapidly limits the number of accessible layers: since only one layer can be exactly corrected for (probably in the middle of the disk), the spot-size is maximum at either the top or bottom layers. Birefringence of the substrate should also become more problematic as the optical path lengths inside the material increase. Despite these difficulties, early researchers working on multilayer optical storage believed that the number of layers could be increased to 20 simply by increasing the read power. For read–write disks, more thought is involved because the read power must remain below the write threshold, and the absorption and recording characteristics as well as the reflectivity of the written state must be individually tuned for each layer. However, by choosing the transmissivity of each layer carefully, a design point of 20 writable layers with a 100mW laser was found. (Note that the media design for just a two layer recordable disk is already quite involved.

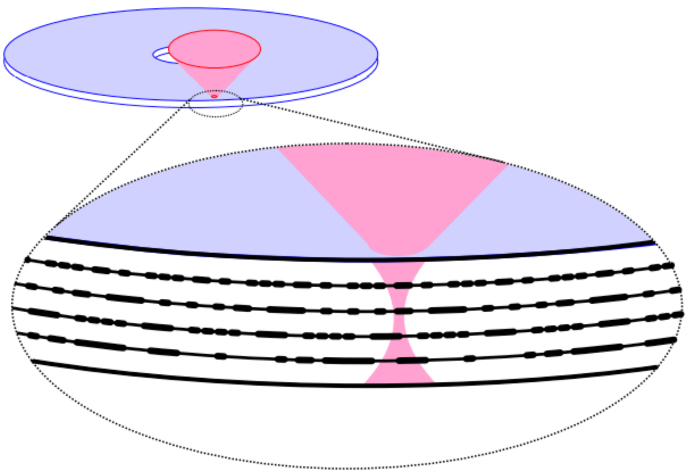


Figure 2. A pre–fabricated disk with multiple layers is accessed from one surface for bit–localized volumetric optical storage.

An additional and important real–world consideration to extending a next–generation–DVD standard beyond two layers is the difficulty of fabricating prerecorded disks with more than two read layers. Currently, while DVD disks can contain as many as four layers, only two layers are read from any surface (i.e., the four–layer disk must be flipped over in a standard one–head drive). So in fabrication, each half of the disk is fabricated with injection molding and is then glued to another half to produce the final disk. If a second data layer is needed on either half, then a 55 micron (±15 micron) thick layer of UV curable polymer is laid down and the bottom layer stamped into this. Apparently, one of the main headaches is gluing the substrates together without warping the resulting disk, since commercial DVD drives do not contain a tilt servo to compensate for warped disks. Other problems include registering the layers (with same center of rotation) so that a read/write head that is tracking on one layer can move to another and still know which track it will be on, and keeping the layer thicknesses uniform so that the focus servo can always distinguish between layers.

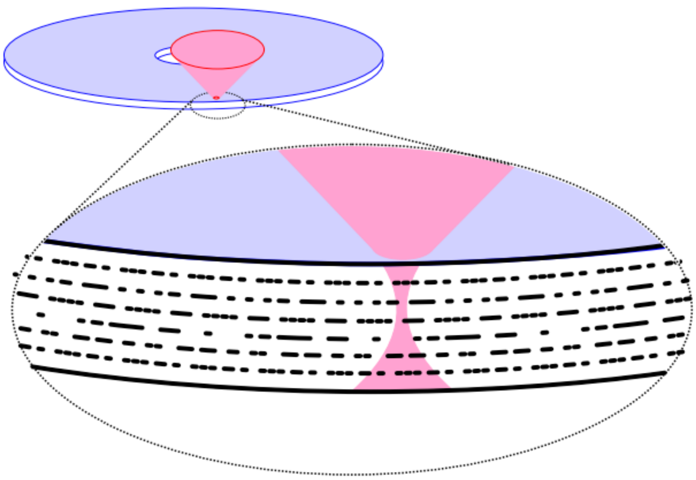
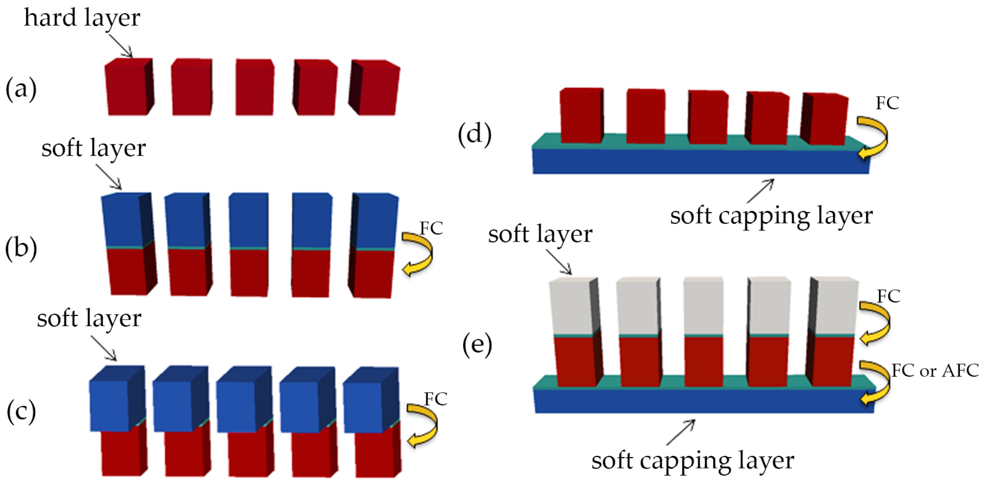


Figure 3. Bits are localized within a initially homogeneous block of media by careful focusing and either confocal imaging, nonlinear material response, or both.

Bit Layers in Homogeneous Media

In contrast with these pre–layered disk schemes, much of the recent scientific literature on localized bit–serial storage has considered the storage of bit layers within initially homogeneous media (Figure 3). Localized–bit volumetric optical data storage has been demonstrated using two–photon fluorescence, by bleaching two–photon fluorescence, by refractive changes in photochromic, photopolymer, photorefractive, and photorefractive polymer materials, and by generating micro-explosions in glass through two–photon absorption. Exposure times have ranged from seconds for photorefractives down to single 100 femtosecond pulses for glass.



With large refractive index changes or 2–photon fluorescence, the written data pattern can be read with an ordinary microscope. Otherwise, a confocal microscope is usually required to read the data with the desired depth selectivity. The first confocal experiments were performed with a differential interference contrast microscope (as described above). For the more widely–used transmission phase–contrast confocal microscope, it turns out that the optical transfer function does not always provide enough spatial frequency coverage to adequately distinguish localized bits, depending on how the bits were recorded. Essentially, such a microscope cannot “see” un-slanted reflection gratings of any grating pitch. Despite this, several early experiments were performed with special annular optics to prepare the readout focus, or a split detector to perform differential measurements. A reflection confocal microscope has advantages over the transmission microscope, because the background signal due to scatter and local index inhomogeneities is greatly reduced. Having all the optics located on one side of the sample also reduces complexity, and allows a single adjustment of microscope tube length to compensate for spherical aberration. Although a reflection phase– contrast microscope has problems similar to the transmission version in covering the spatial frequency band, if the numerical aperture (N.A.) is increased sufficiently and the writing and readout wavelengths chosen carefully, localized bits can be selectively detected. Several nice results were obtained this way, but unfortunately, the need to have an oil–immersion objective (for N.A.>1.0) limits the general applicability.

Thus research in localized–bit volumetric recording appears to have settled on two–photon processes and fluorescence. This includes one–photon confocal detection of fluorescence changes induced by two–photon absorption, and detection with an ordinary microscope of two–photon fluorescence in optically–ablated silica. In the former, high intensity CW illumination can also be used to read and write the data; in the latter, the ablated bits can also be detected by their large refractive index change. Since the photoluminescence, whose origins in optically damaged glasses is not yet well understood, can be erased by annealing, researchers hope to store two bits per voxel through this method. In addition, by sectioning the glass and measuring the exposed voxels with an atomic–force microscope, Glezer and Mazur found the written marks to be significantly smaller than the optical spot size, an effect attributed to self–focusing of the sub–picosecond laser pulse.

Bit–Parallel Readout of 2–Photon Fluorescence

Moving from bit–serial to bit–parallel access is an attractive way to increase the potential data transfer rates. With two–photon fluorescence readout, a data–set containing many bits can be selected and read out in parallel. A number of papers on bit–parallel two–photon memories have been produced, starting from the initial materials work by Parthenopoulos and Rentzepis and subsequent systems efforts of Esener and co–workers. In this scheme, the two–photon process is selectively applied to a localized region or spot by applying two beams through orthogonal faces of a cube of material. For bit–serial access, both beams are focused ; for bit–parallel access, one beam contains a page of information imaged to a plane within the material while the other beam is a cylindrically focused sheet of light illuminating only this image plane from the side (See Figure 4).

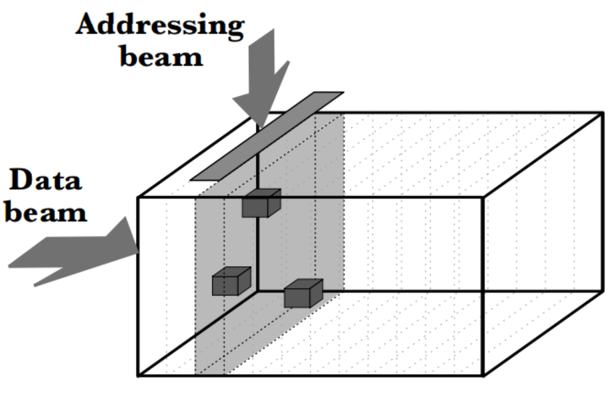


Figure 4. Orthogonal beams used to write and read data in parallel in 3–dimensions using 2–photon fluorescence.

Initially, spirobenzopyran embedded in polymer was used because its desirable photochromic behavior (that is, light can induce a change in the molecule's absorption spectrum) was accessible with the fundamental and second harmonic of Nd:YAG lasers. The energy levels are shown in Figure 5 for the unwritten and written forms of the molecule. The orthogonal arrangement of beams is required because of the potential for two–photon absorption by the green beam alone. By using the green light as the cylindrical addressing beam, the two–photon write process is confined to this illuminated 2–D plane. As shown in Figure 5, written layers can be illuminated either by one– or two–photon fluorescence. Using one–photon fluorescence (green addressing beam for readout) offers more efficiency and does not necessarily imply destructive readout. The orthogonal arrangement also provides the opportunity for multi–functional access of database records using the different faces of the storage cube for different database actions. After these initial studies, subsequent materials development then attempted to improve the response of spiropyrans or to find other suitable photochromic materials. A list of desired characteristics, adapted from Reference, might include:

- a high two–photon absorption cross–section so that incident light is used efficiently to write information, and/or high doping levels to increase the number of molecules per voxel;
- the written form of the molecule should have a high fluorescence quantum yield, for sufficient readout signal strength. Note that since the fluorescence is emitted in all directions, only the portion captured by the numerical aperture of the readout optics contributes to the detected signal strength (One proposal for improving this is to use 3–D photonic bandgap structures at each voxel in order to generate fluorescence only in the desired direction);
- both forms (written and unwritten) of the photochromic molecule should be stable at (and even above) room temperature;
- high fatigue resistance (i.e., few residual products in the photochemical reactions) for > 106 write–read–erase cycles;
- a large separation in absorption and emission spectra so that readout signals can be filtered out, and so that the fluorescence signal does not affect either written or unwritten molecules;
- nondestructive readout (also linked to the spectral separations);
- and the capability of being fabricated in low–scatter, high–optical quality samples using simple polymer hosts.

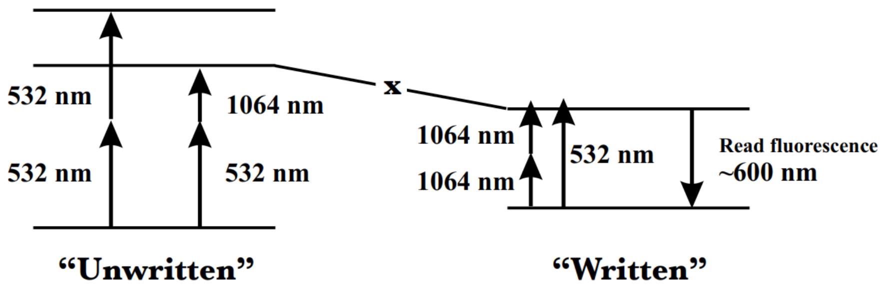
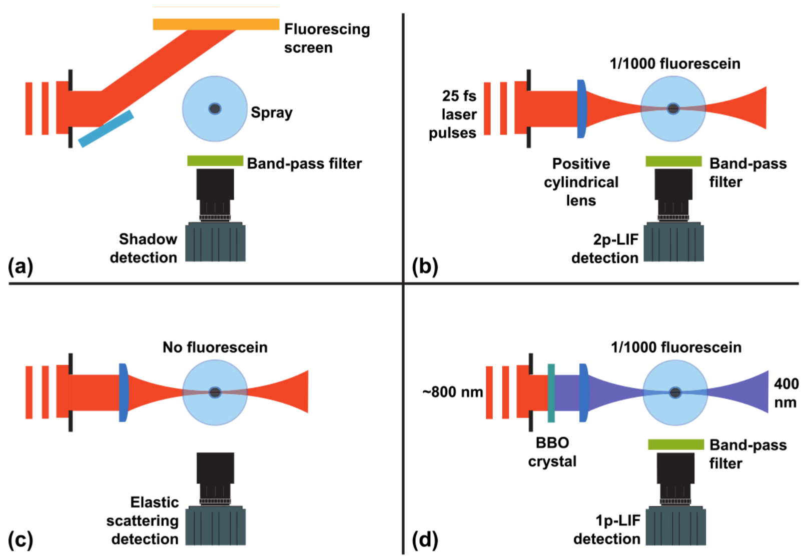


Figure 5. Energy band diagrams for the unwritten and written forms of the fluorescent photochromic material spirobenzopyran.

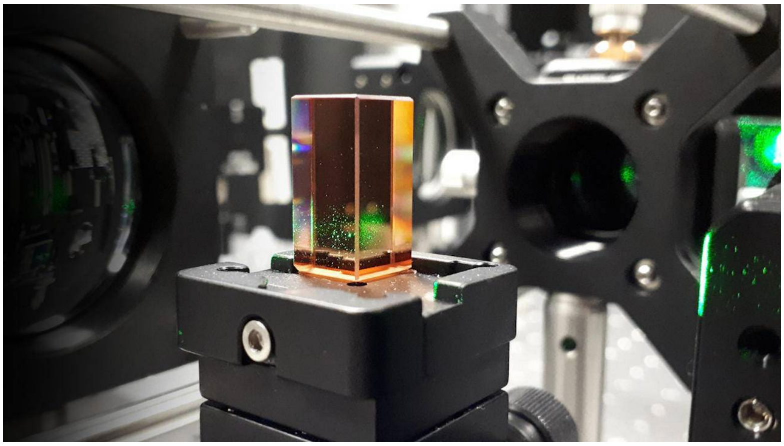
In addition to these rewritable photochromics, write–once materials, composed of organic dyes triggered by photoacid generators, have also been developed. For rapid generation of pre–recorded disks, a "stamping" scheme was proposed, using a volume holographic element read with multiple mutually incoherent reference beams to generate the recording optical signals at multiple layers simultaneously. Some of the systems difficulties with this two–photon parallel–access method stem from the media. The low sensitivity of the two–photon process requires high–power pulsed lasers, which means that any optics with an intermediate focus must be enclosed in vacuum, and the optics and the spatial light modulator for imposing pixelated data patterns must have high damage thresholds. Despite the high powers, the media still requires hundreds of recording pulses per data plane, or forces serial access during writing.



Other difficulties are inherent to the system: the need to have a sheet of light implies that the effective width of each data layer may be much larger than the wavelength. Marhic pointed out that this tradeoff between density and transfer rate can actually limit the effective areal density to the same $1/\lambda 2$ limit as surface optical storage. Another difficulty is the need to refocus the output detector array onto each data plane to be read. This problem can be solved by arranging the data planes in a "turbofan" arrangement, so that each data plane contains the radius of a spinning cylinder or thick disk. In this way, the disk rotation brings each data plane to exactly the same position relative to an external read head. The readout light enters the surface of the disk illuminating a tilted plane, and the fluorescent signals are detected back through the top of the disk, using anamorphic optics to correct for the change in path length between the disk entrance face and the different points on the data plane. Using such a scheme, twenty–five layers have been recorded and read with a layer spacing of 75 microns. More recently, a system capable of being scaled up to faster readout per channel was designed for four layers each with 16 data bits.

Holographic Storage

In contrast to localized–bit recording, where each bit of data is assigned to a particular location within the storage volume, holographic storage distributes data throughout a volume in a delocalized way. A hologram is a recording of the optical interference pattern that forms at the intersection of two coherent optical beams (object and reference— Figure 6(a)). The object beam carries the information to be stored, while the reference beam is designed to be simple to reproduce at a later stage. (A common reference beam is a plane wave: a light beam that propagates without converging or diverging.)



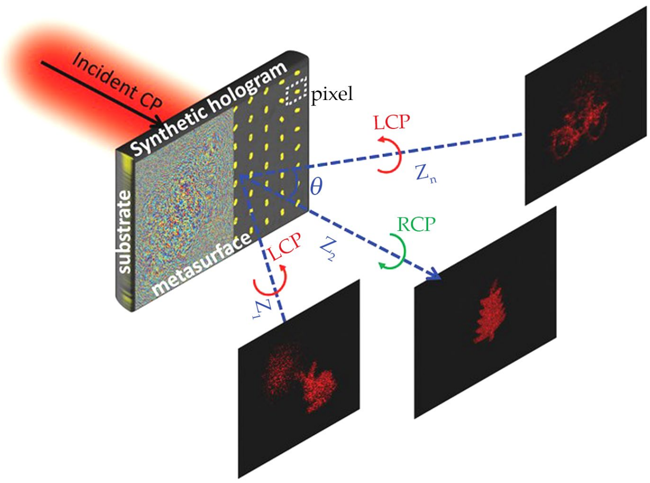
To record a hologram, the reference and object beams are made to overlap in a photosensitive medium, such as a photopolymer or inorganic crystal or even photographic film, where the resulting optical interference pattern creates chemical and/or physical changes. As a result, a replica of the interference pattern is stored as a change in absorption, refractive index or thickness. Since the pattern contains information about both the amplitude and the phase of the two light beams, when the recording is illuminated by the readout beam, some of the light is diffracted to "reconstruct" a weak copy of the object beam (Figure 6(b)). If the object beam originally came from a 3–D object, then the reconstructed hologram makes the 3–D object reappear. Although holography was conceived in the late 1940s, it was not considered a potential storage technology until the development of the laser in the 1960s. The resulting rapid development of holography for displaying 3–D images led researchers to realize that potentially, holograms could also store data at a volumetric density of $1/\lambda^3$.

In holographic storage, data sets are transferred to and from the storage material as 2–D images composed of thousands of pixels, with each pixel representing a single bit of information. However, no one location in the crystal is responsible for storing that one bit; each bit is distributed throughout the recorded interference fringes. Since an entire "page of data" can be retrieved by a photodetector at the same time, rather than bit-by-bit, the holographic scheme promises fast readout rates as well as high density. If a thousand holograms, each containing a million pixels, could be retrieved every second, for instance, then the output data rate would reach 1 Gigabit per second. Despite this attractive potential and fairly impressive early progress, however, research into holographic data storage all but died out in the mid-1970s mostly because of the lack of suitable devices for the input and output of pixelated 2–D data pages.

In the early 1990s, interest in volume-holographic data storage was rekindled by the availability of devices that could display and detect 2–D pages, including charge coupled devices (CCD), complementary metal-oxide semiconductor (CMOS) detector chips and small liquid-crystal panels. The wide availability of these devices was made possible by the commercial success of hand-held camcorders, digital cameras, and video projectors. With these components in hand, holographic-storage researchers have begun to demonstrate the potential of this technology in the laboratory. By using the volume of the media, researchers have experimentally demonstrated that data can be stored at equivalent areal densities of nearly 400 bits/sq. micron. (For comparison, a single–layer of a DVD disk stores data at ~ 4.7 bits/sq. micron.) A readout rate of 10 Gigabit per second has also been achieved in the laboratory.

Holographic Multiplexing

If the hologram is recorded in a thin material—such as the security hologram stamped onto many credit cards—the readout beam can differ in angle or wavelength from the reference beam used for recording the image. The scene will still appear. However, if the hologram is recorded in a thick material, the reconstructed object beam will only appear when the readout beam is nearly identical to the original reference beam. Since the diffracted wavefront accumulates energy from throughout the thickness of the storage material, a small change in either the wavelength or angle of the readout beam generates enough destructive interference to make the reconstructed object beam effectively disappear.



As the material becomes thicker, accessing a stored volume hologram requires tight tolerances on the stability and repeatability of the wavelength and the angle provided by the laser and readout optics. However, destructive interference also opens up a tremendous opportunity: a small storage volume can now store multiple superimposed holograms, each one distributed throughout the entire volume. The destructive interference allows each of these stored holograms to be selectively accessed with its original reference beam. Several different techniques have been developed to define a set of suitable reference beams by, for example, slightly changing the angle, wavelength or phase–front of the original light beam. Using so-called "angle multiplexing," as many as 10,000 holograms have been stored in a 1 cm^3 volume.

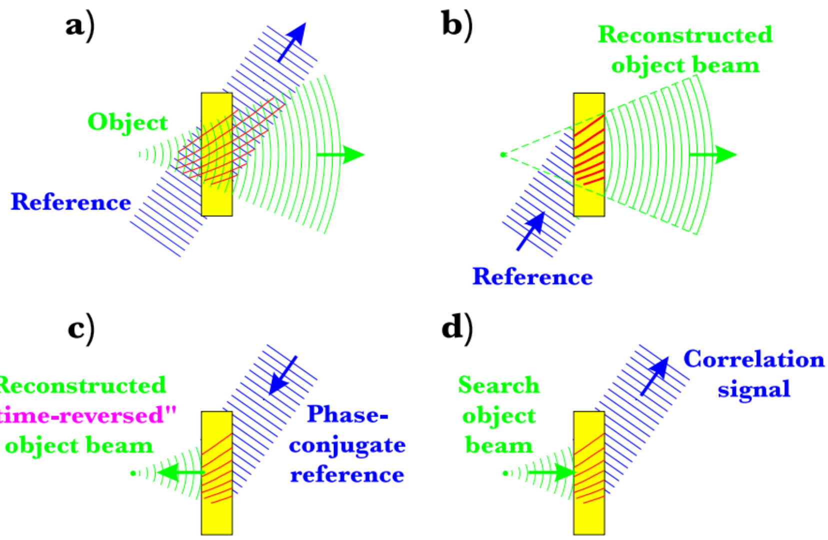


Figure 6. How to record and read data using holograms: (a) Holographic storage of a single data bit. The spherical wave from a single pixel interferes with a coherent plane wave in the reference beam. The resulting interference pattern changes the refractive properties of the photosensitive medium. (b) The hologram is read out using the original reference beam, which is diffracted by the stored interference pattern to reconstruct the original spherical wavefront. An image of this beam can be formed on a single detector pixel, resulting in the retrieval of a single bit. (c) The hologram can also be read out by illuminating it with a counter-propagating (or "phase-conjugate") reference beam, which reconstructs a phase-conjugate copy of the original object beam. This beam returns to its original point of origin, where the bit value can be read without requiring a high-quality imaging system. (d) A third way to retrieve data involves illuminating it with a diverging object beam, which reconstructs the original plane wave reference beam. This beam can be focused onto a detector and provides an optical measurement of the correlation between the stored data and the illuminating object beam. This technique can allow one to search the stored data according to its content, rather than according to its address.

1.1.5 Advantages and Disadvantages of 3D Optical Data Storage

Optical storage varies from other data storage techniques that make use of other technologies such as magnetism or semiconductors.

Upon using 3D optical data storage:

- Optical media can last a long time depending on what kind of optical media you choose but only with proper care.
- It is great for archiving. Meaning, when data is written to them, they cannot be reused – the data is permanently preserved having no possibility of being overwritten.
- It is widely used on other platforms (PCs or any other system).
- Optical media has the capability in pinpointing a particular piece of data stored on it, independent of the other data on the volume or the other in which that data was stored on the volume.

The downsides can be:

- Due to system's write-once-read-many (WORM) characteristic, it prevents you from being able to use that media again.
- The server employs software compression to write compressed data to your optical media taking considerable processing unit sources. Thus, increasing the time needed in writing and restoring that data.

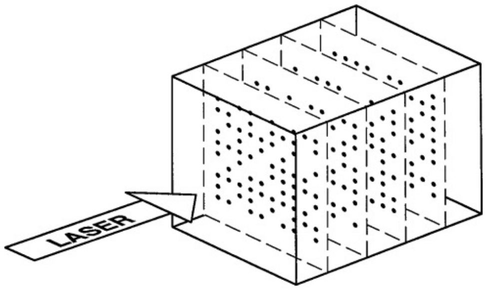
Although several companies are actively developing the technology and claiming that it may become available soon, still no commercial product based on 3D optical data storage has yet arrived due to design issues that need to be addressed.

1.2 BASIC COMPONENTS IN 3-D OPTICAL DATA STORAGE

The process of storing and retrieving data in optical devices requires special devices. The following involves a discussion of some of these components.

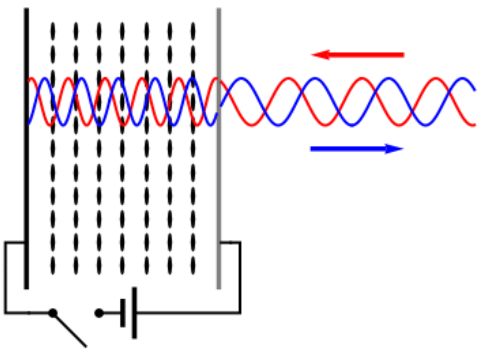
1.2.1 Laser

Laser (light amplification by stimulated emission of radiation) is a device used to generate coherent and almost monochromatic radiation. The super directional electromagnetic radiation is emitted somewhere in the ranges of submillimeter through X-ray and ultraviolet. The fundamental attributes of a laser beam that make them ideal for optical recording are their directionality, coherence, monochromaticity, and brightness. Krypton lasers, argon ion lasers, and diode lasers are used to record holograms on crystals.



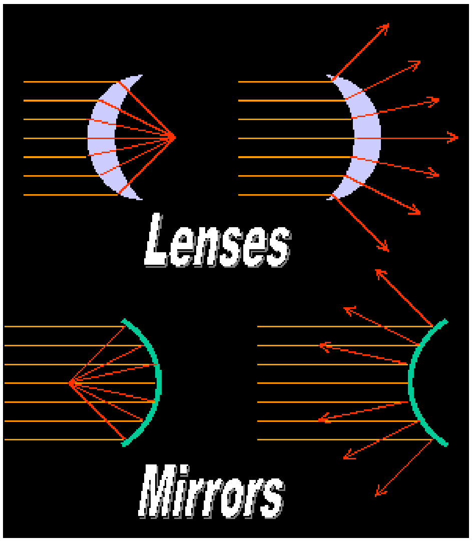
1.2.2 Spatial Light Modulators (SLM)

During recording, the real image or data is converted to a single beam of light which intersects with the reference beam. This conversion is done by a spatial light modulator device. The SLM consists of an array of pixels that are usually composed of microscopic shatter that can be controlled by a computer. LCD displays are also used in place of microscopic shutters. Ever SLM pixel corresponds to a bit of data. Therefore, the bits cause a shut or open in microscopic shutters or go dark or transparent depending on the condition of the bit whether it 0 or 1.



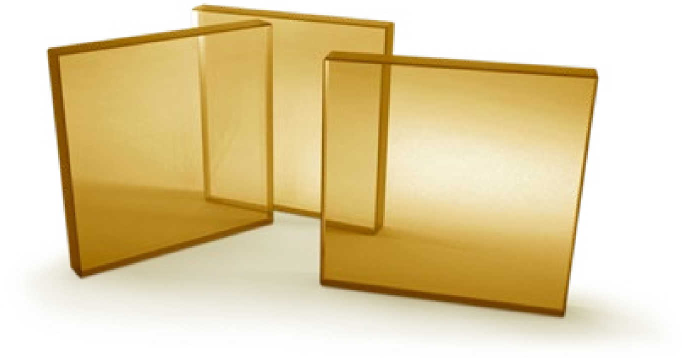
1.2.3 Lens and Mirrors

The lenses and mirrors are the major components used to achieve an inverse transform and a Fourier transforms optically. To form the transform a collimated beam projected on transparency and the transform lens makes the parallel ray bundles make a convergence of the focal plane of the lens. The back focal plane, referred to as the Fourier transform plane, is where the transformation of the spatial image is done into a spatial frequency spectrum. The lens carries out, at the speed of light a two dimensional Fourier transform, which leads to the formation of a far-field diffraction pattern that can be observed on a screen placed on the transform plane. The square of Fourier transform amplitude of the input signal determines the intensity of the pattern. A spatial frequency lobe of the image is removed using a stop placed at particular frequency lobe on this plane. Two lenses are used in the processes, with the aperture of the lens limiting the Fourier transform resolution and the second lens forming the inverse transform as well as recovering the original signal.



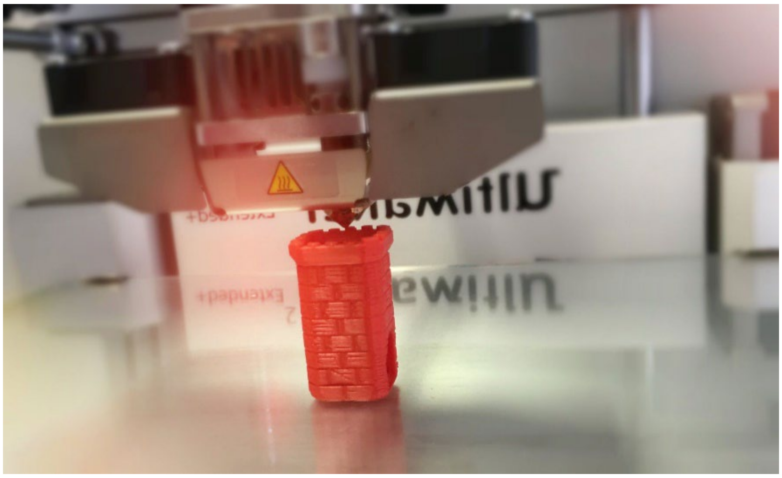
1.2.4 Photorefractive Crystals

Photopolymers and photorefractive crystals are the two main materials used in the holographic storage devices. To record data, photorefractive crystals of material such as BaTiO3 or LiNbO3 are used due to their optical properties such as high diffraction efficiency and high resolution. Additionally, the materials also show the ability to store data permanently till erasure as well as fast erasure through the application of an external stimulus such as ultraviolet light.



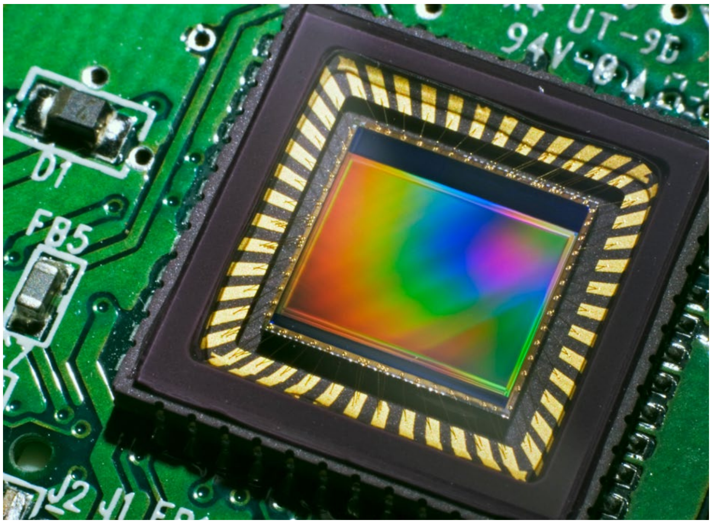
1.2.5 Photopolymers

The recent advancement in technology is in the development of photopolymers that can be used as a medium of holographic storages. Photopolymers have less thickness compared to photorefractive crystals as polymers are usually limited by their optical quality and mechanical stability. Some of the photopolymers used in the market include HRF-150 from DuPont. This film from DuPont can reach 12bits/ μm^2 when used with a thickness of 100 millimeters which is much greater than DVD-ROM with a factor of 2. The imprinting of the pattern in the photopolymer is done by inducing photochemical changes.



1.2.6 Charge Coupled Devices (CCD)

Charge-coupled devices are silicon-based devices that contain an array of wells and are created through implants and a series of column. They provide the mechanism to convert optical images to electrical signals.



1.3 3D OPTICAL STORAGE DATA WRITING PROCESSES

Data recording in a 3D optical storage medium requires that a change take place in the medium upon excitation. This change is generally a photochemical reaction of some sort, although other possibilities exist. Chemical reactions that have been investigated include photoisomerizations, photodecompositions and photobleaching, and polymerization initiation. Most investigated have been photochromic compounds, which include azobenzenes, spiropyrans, stilbenes, fulgides, and diarylethenes. If the photochemical change is reversible, then rewritable data storage may be achieved, at least in principle. Also, MultiLevel Recording, where data is written in "grayscale" rather than as "on" and "off" signals, is technically feasible.

1.3.1 Writing by Nonresonant Multiphoton Absorption

Although there are many nonlinear optical phenomena, only multiphoton absorption is capable of injecting into the media the significant energy required to electronically excite molecular species and cause chemical reactions. Two-photon absorption is the strongest multiphoton absorbance by far, but still it is a very weak phenomenon, leading to low media sensitivity. Therefore, much research has been directed at providing chromophores with high two-photon absorption cross-sections.

Writing by two-photon absorption can be achieved by focusing the writing laser on the point where the photochemical writing process is required. The wavelength of the writing laser is chosen such that it is not linearly absorbed by the medium, and therefore it does not interact with the medium except at the focal point. At the focal point two-photon absorption becomes significant, because it is a nonlinear process dependent on the square of the laser fluence.

Writing by two-photon absorption can also be achieved by the action of two lasers in coincidence. This method is typically used to achieve the parallel writing of information at once. One laser passes through the media, defining a line or plane. The second laser is then directed at the points on that line or plane that writing is desired. The coincidence of the lasers at these points excited two-photon absorption, leading to writing photochemistry.

1.3.2 Writing by Sequential Multiphoton Absorption

Another approach to improving media sensitivity has been to employ resonant two-photon absorption (also known as "1+1" or "sequential" two–photon absorbance). Nonresonant two-photon absorption (as is generally used) is weak since in order for excitation to take place, the two exciting photons must arrive at the chromophore at almost exactly the same time. This is because the chromophore is unable to interact with a single photon alone. However, if the chromophore has an energy level corresponding to the (weak) absorption of one photon then this may be used as a stepping stone, allowing more freedom in the arrival time of photons and therefore a much higher sensitivity. However, this approach results in a loss of nonlinearity compared to nonresonant two–photon absorbance (since each two-photon absorption step is essentially linear), and therefore risks compromising the 3D resolution of the system.

1.3.3 Microholography

In microholography, focused beams of light are used to record submicrometre-sized holograms in a photorefractive material, usually by the use of collinear beams. The writing process may use the same kinds of media that are used in other types of holographic data storage, and may use two–photon processes to form the holograms.

1.3.4 Data Recording during Manufacturing

Data may also be created in the manufacturing of the media, as is the case with most optical disc formats for commercial data distribution. In this case, the user cannot write to the disc – it is a ROM format. Data may be written by a nonlinear optical method, but in this case the use of very high power lasers is acceptable so media sensitivity becomes less of an issue.

The fabrication of discs containing data molded or printed into their 3D structure has also been demonstrated. For example, a disc containing data in 3D may be constructed by sandwiching together a large number of wafer-thin discs, each of which is molded or printed with a single layer of information. The resulting ROM disc can then be read using a 3D reading method.

1.3.5 Other Approaches to Writing

Other techniques for writing data in three-dimensions have also been examined, including:

Persistent spectral hole burning (PSHB), which also allows the possibility of spectral multiplexing to increase data density. However, PSHB media currently requires extremely low temperatures to be maintained in order to avoid data loss.

Void formation, where microscopic bubbles are introduced into a media by high intensity laser irradiation.

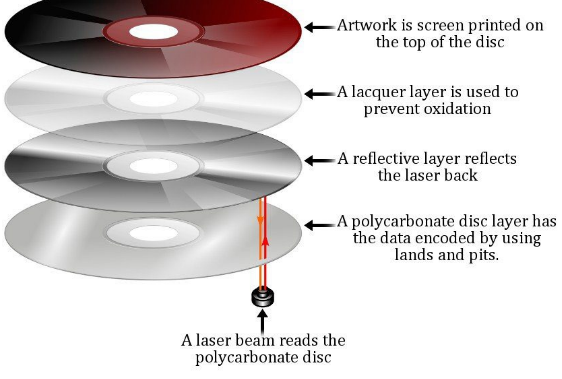
Chromophore poling, where the laser-induced reorientation of chromophores in the media structure leads to readable changes.

1.4 3D OPTICAL DATA STORAGE READING PROCESSES

The reading of data from 3D optical memories has been carried out in many different ways. While some of these rely on the nonlinearity of the light-matter interaction to obtain 3D resolution, others use methods that spatially filter the media's linear response. Reading methods include:

Two photon absorption (resulting in either absorption or fluorescence). This method is essentially two-photon microscopy.

Linear excitation of fluorescence with confocal detection. This method is essentially confocal laser scanning microscopy. It offers excitation with much lower laser powers than does two-photon absorbance, but has some potential problems because the addressing light interacts with many other data points in addition to the one being addressed.



Measurement of small differences in the refractive index between the two data states. This method usually employs a phase contrast microscope or confocal reflection microscope. No absorption of light is necessary, so there is no risk of damaging data while reading, but the required refractive index mismatch in the disc may limit the thickness (i.e. number of data layers) that the media can reach due to the accumulated random wavefront errors that destroy the focused spot quality.

Second harmonic generation has been demonstrated as a method to read data written into a poled polymer matrix.

Optical coherence tomography has also been demonstrated as a parallel reading method.

1.5 MEDIA DESIGN

The active part of 3D optical storage media is usually an organic polymer either doped or grafted with the photochemically active species. Alternatively, crystalline and sol-gel materials have been used.

1.5.1 Media form factor

Media for 3D optical data storage have been suggested in several form factors: disk, card and crystal.

A disc media offers a progression from CD/DVD, and allows reading and writing to be carried out by the familiar spinning disc method.

A credit card form factor media is attractive from the point of view of portability and convenience, but would be of a lower capacity than a disc.

Several science fiction writers have suggested small solids that store massive amounts of information, and at least in principle this could be achieved with 5D optical data storage.

1.5.2 Media manufacturing

The simplest method of manufacturing – the molding of a disk in one piece – is a possibility for some systems. A more complex method of media manufacturing is for the media to be constructed layer by layer. This is required if the data is to be physically created during manufacture. However, layer-by-layer construction need not mean the sandwiching of many layers together. Another alternative is to create the medium in a form analogous to a roll of adhesive tape.

1.6 DEVELOPMENT ISSUES

Despite the highly attractive nature of 3D optical data storage, the development of commercial products has taken a significant length of time. This results from limited financial backing in the field, as well as technical issues, including:

1.6.1 Destructive reading

Since both the reading and the writing of data are carried out with laser beams, there is a potential for the reading process to cause a small amount of writing. In this case, the repeated reading of data may eventually serve to erase it (this also happens in phase change materials used in some DVDs). This issue has been addressed by many approaches, such as the use of different absorption bands for each process (reading and writing), or the use of a reading method that does not involve the absorption of energy.

1.6.2 Thermodynamic stability

Many chemical reactions that appear not to take place in fact happen very slowly. In addition, many reactions that appear to have happened can slowly reverse themselves. Since most 3D media are based on chemical reactions, there is therefore a risk that either the unwritten points will slowly become written or that the written points will slowly revert to being unwritten. This issue is particularly serious for the spiropyrans, but extensive research was conducted to find more stable chromophores for 3D memories.

1.6.3 Media sensitivity

Two-photon absorption is a weak phenomenon, and therefore high power lasers are usually required to produce it. Researchers typically use Ti-sapphire lasers or Nd:YAG lasers to achieve excitation, but these instruments are not suitable for use in consumer products.

SUMMARY

- 3D optical data storage is the term given to any form of optical data storage in which information can be recorded and/or read with three dimensional-resolution (as opposed to the two dimensional resolution afforded, for example, by CD).
- Different from 2Ddata storage where data is stored on an aluminum film built in a disk and read through focusing of laser beams on its surface, 3D optical data storage uses nonlinear methods of reading and recording data.
- Optical storage systems consist of a drive unit and a storage medium in a rotating disk form. In general the disks are pre-formatted using grooves and lands (tracks) to enable the positioning of an optical pick-up and recording head to access the information on the disk.
- 3D optical data storage is related to (and competes with) holographic data storage.
- The process of storing and retrieving data in optical devices requires special devices.
- Data recording in a 3D optical storage medium requires that a change take place in the medium upon excitation. This change is generally a photochemical reaction of some sort, although other possibilities exist.
- The reading of data from 3D optical memories has been carried out in many different ways. While some of these rely on the nonlinearity of the light-matter interaction to obtain 3D resolution, others use methods that spatially filter the media's linear response.
- Despite the highly attractive nature of 3D optical data storage, the development of commercial products has taken a significant length of time.

REFERENCES

1. Burr, G.W. (2003). Three-Dimensional Optical Storage (PDF). SPIE Conference on Nano-and Micro-Optics for Information Systems. pp. 5225–16. Archived from the original (PDF) on March 8, 2008.
2. Christenson, C. W., Saini, A., & Singer, K. D. (2016). 3D Optical Data Storage in Organic Film using Single Photon Absorption. Optical Data Processing and Storage, 2(1).
3. Fort, A. F.; Barsella, A.; Boeglin, A. J.; Mager, L.; Gindre, D.; Dorkenoo, K. D. (29 August 2007). Optical storage through second harmonic signals in organic films. SPIE Optics+Photonics. San Diego, USA. pp. 6653–10.
4. G. W. Burr, "Volumetric storage," in Encyclopedia of optical engineering, R. B. Johnson and R. G. Driggers, eds., Marcel Dekker, New York, 2003.
5. G. W. Burr, C. M. Jefferson, H. Coufal, M. Jurich, J. A. Hoffnagle, R. M. Macfarlane, and R. M. Shelby, "Volume holographic data storage at areal density of 250 gigapixels/in2 ," 26(7), pp. 444–446, 2001.
6. Kawata, S., & Kawata, Y. (2000). Three-dimensional optical data storage using photochromic materials. Chemical reviews, 100(5), 1777-1788.
7. Kawata, S.; Kawata, Y. (2000). "Three-Dimensional Optical Data Storage Using Photochromic Materials". Chemical Reviews. 100 (5): 1777–88. doi:10.1021/cr980073p. PMID 11777420.
8. T. Imai, S. Yagi, A. Tate, T. Kurokawa, T. Tanabe, M. Yamamoto, and S. Imamura, "Noise analysis and reduction of waveguide holographic read–only memory," in Annual Meeting of the Optical Society of America, p. 59, Oct 2000.

CHAPTER 2

OPTICAL STORAGE

INTRODUCTION

Optical storage is the storage of data on an optically readable medium. Data is recorded by making marks in a pattern that can be read back with the aid of light, usually a beam of laser light precisely focused on a spinning optical disc. An older example of optical storage that does not require the use of computers, is microform. There are other means of optically storing data and new methods are in development. An optical disc drive is a device in a computer that can read CD-ROMs or other optical discs, such as DVDs and Blu-ray discs. Optical storage differs from other data storage techniques that make use of other technologies such as magnetism, such as floppy disks and hard disks, or semiconductors, such as flash memory.

Optical storage is any storage type in which data is written and read with a laser. Typically, data is written to optical media such as compact discs (CDs) and digital versatile discs (DVDs). At one time, optical discs were considered a potential replacement for hard disk drives (HDDs) in computing systems, but their lack of growth in capacity compared to both HDDs and later flash-based solid-state drives (SSDs) has relegated optical storage use mostly to long-term archiving and data backup.

Although optical media is more durable and less vulnerable to environmental conditions than tape, HDDs and SSDs, optical discs are slower than the typical HDD and significantly slower than the SSD and offer lower storage capacities than either. Blu-ray disks are currently the fastest optical media on the market and provide much more capacity than CDs and DVDs, but they still lag behind HDDs and SSDs.

2.1 OPTICAL STORAGE SYSTEM

Use the information that is described as an overview and reference guide for IBM® optical support to a system with the IBM i operating system. *Optical storage* is any storage method that uses a laser to store and retrieve data from optical media.

Examples of this media are compact disk read-only memory (CD-ROM), digital versatile disk read-only memory (DVD-ROM), digital versatile disk random access memory (DVD-RAM), write-once read-many (WORM) cartridges, erasable optical cartridges, and Removable Mass Storage (RMS) media which are removable disk (RDX) and flash drives.

These functions are unique to optical support:

- CD-ROM devices
- DVD devices
- Directly attached optical media library devices
- LAN-attached optical media library devices
- Virtual optical devices
- RDX
- Flash drives

RMS devices and media are considered optical storage class for the purposes of supporting commands and functions on IBM i. All references to DVD devices and media in the optical storage section can be considered to also include RMS devices and media. For more information about RMS media and devices.

This information is intended for the following users:

- System operators and users can use this information as their primary reference for CD-ROM, DVD, optical media libraries, and virtual optical support.
- Service representatives can use this information to perform activities as directed by the appropriate optical device service guides.

Optical storage on the system provides an economical and efficient way to store and retrieve large amounts of information at a high performance level. Optical storage devices offer significant advantages over other high-capacity storages devices, such as tape and microfilm, with faster access times and a hierarchical-type file organization. IBM i optical storage uses files that are stored in directories and files that are stored in subdirectories similar to UNIX or PC-based file systems.

The capacity, price, and performance of optical storage continually improve, and IBM remains committed to providing its customers with these improvements over time. Even as new devices are introduced, the basic methods of accessing optical information remain consistent, as these new storage devices are being added under the current file system interfaces that optical storage programs have used for years.

These are some considerations in the use of optical storage media:

Consideration	Reason for use
Durability	Optical media can have a shelf life in excess of 50 years.
Archive storage	Write-once read-many (WORM) optical media can be used to archive large amounts of data. Each sector on the media is only written once when creating and updating files and directories. When a file is changed or deleted, a new version of the file gets written, but the old version still exists on the media. All previous versions of the file remain recorded on the media. This capability also exists on erasable media, but the entire disk can be erased and reused.
Transportability	Universal Disk Format (UDF) optical media can be read with any other industry operating system platform that supports UDF, which is an industry standard file system. Optical Media written with High Performance Optical File System (HPOFS) format can be interchanged with other optical media libraries attached to a system.
Random access	Optical devices are random access devices. This facilitates the retrieval of relevant data on demand. File access is independent of the order in which the data was stored. Also, multiple users can access the same volume at the same time.

When you use virtual optical storage, you create and use optical images that are stored on your disk units. These optical images are treated as if they were real optical disk media by the internal file system functions. The term virtual applies to the emulation of the optical media sectors when used by read and write functions.

2.1.1 Common Examples of Optical Media

Over the years, optical media have taken a variety of forms, including Laser Disc (LD), HD-DVD, write-once, read-many (WORM) optical cartridges and several others. From this mix, three formats have emerged as standards in today's optical storage market:

- **Compact disk (CD).** The CD represents the first generation of commercial optical storage. After its introduction, it quickly replaced both vinyl records and cassette tapes as the audio medium of choice. Originally, CDs were available only as prerecorded read-only disks, but it wasn›t long before they became available as recordable discs and rewritable discs that could be used for data storage. The CD can hold up to up 700 megabytes (MBs) of data.

- **Digital versatile disc (DVD).** Also referred to as the digital video disc, the DVD started out as a read-only medium similar to CDs but with the ability to hold enough data to store a full-length movie. A single-layer DVD can hold 4.7 gigabytes (GB) of data, and a double-layer disc can hold 8.5 GB. Not long after DVDs were introduced, recordable and rewritable discs became available for data storage.
- **Blu-ray.** The Blu-ray disk has emerged as clear leader in today›s optical storage market. Unlike CDs and DVDs, which use a red laser to read and write data, a Blu-ray disk uses a blue laser, which dramatically increases capacities and data transfer rates over CDs and DVDs. Today›s Blu-ray discs can store up to 128 GB of data and are available as read-only disks that can hold prerecorded high-definition feature films as well as recordable and rewritable disks for data storage.

The standard size for all three formats is the same: 120 mm (4.7 inches) in diameter and 1.2 mm (0.05 inches) thick. This standard makes it possible for Blu-ray drives to support DVDs and CDs and for DVD drives to support CDs. That said, optical drives are compatible only with earlier formats and not the other way around. CD drives cannot run DVD or Blu-ray discs, and DVD drives cannot run Blu-ray discs.

2.1.2 Advantages and Disadvantages of Optical Storage

One of the biggest advantages of optical storage over other storage media is durability. Optical discs are not vulnerable to data loss due to power failure like volatile memory, and they're not as subject to wear as non-volatile memory (NVM) such as HDDs and flash SSDs. Optical discs are also much sturdier than magnetic tape, which is the leading archival storage medium.

Another advantage of the optical disc is that the storage medium is inexpensive to manufacture, although costs can vary depending on the type of discs and how they're used. Discs that contain prerecorded material such as audio CDs or Blu-ray movies are made up primarily of aluminum foil and plastic. Manufacturers can easily bulk-produce these kind of discs by using a die and stamp technique that presses the tiny pits into the reflective foil medium in an assembly line process.

Optical discs for data storage have different requirements. Unlike prerecorded discs that are written to only once, optical discs for storage typically need to be rewritable, which requires different recording material. Rewritable discs cannot use a low-cost reflective foil layer like prerecorded discs. Instead, they must use a more expensive layer of phase-change material that enables the data to be erased and written over multiple times.

Despite these difference, optical discs for storage are still made up mostly of plastic and can be produced in bulk, making them cheaper to manufacture compared to other storage media.

The biggest disadvantage of optical storage is disk capacity. The latest 12-cm Blu-ray discs top out at 128 GB, far below what is now possible with either HDDs or SSDs on a per-centimeter basis. At the same time, the rise of internet streaming and Universal Serial Bus (USB) flash drives has also diminished the reliance on optical discs.

However, the future of optical storage might not lie with plastic disks, but with quartz crystal, at least for archival or write-once data. Microsoft's Project Silica is actively working on a storage technology that uses ultrafast laser optics to store data in silica glass.

2.1.3 Supported Hardware for Optical Storage

Various stand-alone optical devices are available as optical storage.

A variety of hardware configurations for CD-ROM, DVD-ROM, DVD-RAM, RDX and Flash drives are supported on a system. The table lists the stand-alone optical devices available. To see the supported media capability of a device, enter Display Device Description (DSPDEVD) on the command line.

Table 1. Supported stand-alone optical type devices

Device type	**Hardware resource type and model**	**Device**
632A	632A-005	Standalone USB DVD drive
632B	Virtual device backed up by the integrated file system (632B-001) or network file system (632B-003), or contained in a RMS device (632B-011)	Virtual device
632C	632C-002	Virtual device hosted by another partition.

6320/6321	6320-002/6321-002	CD-ROM
6330 HH DVD-RAM	6330-002	DVD-RAM
6331 Slim DVD RAM	6331-0xx	Slim multi-recorder
6333 HH DVD RAM	6333-002	HH multi-recorder
6336 HH DVD-ROM	6336-002	DVD-ROM
6337 Slim Line DVD-ROM	6337-00x	DVD-ROM
63B8	63B8-004 SATA RDX drive 63B8-005 USB RDX drive 63B8-0D2 RDX drive hosted by another partition	Removable disk (RDX) dock.
63BC	63BC-005 USB flash drive 63BC-0D2 USB flash drive hosted by another partition	Removable Mass Storage (RMS) device that is not an RDX device. A USB flash drive is an example of this type of device.
7210-020	6321-002	CD-ROM Bridgebox external device
7210-025	6330-002	DVD-RAM Bridgebox external device
7210-030	6333-002	External device
7212-102	6330 6333 6336	
7214	6331 6337	Storage device enclosure
7226	6331 or 63B8	Storage device enclosure with DVD-RAM or RDX feature.
5720	6331 6337	Storage device enclosure

The 632C-002 client virtual optical device now supports an embedded media changer when the virtual storage server partition is an IBM® i. For example, if the backing optical device is a 632B-001 or 632B-003, then the client 632C automatically changes media for installation and save restore operations. For enablement and usage information, see the optical storage website or review the latest PSP information.

Optical media libraries come in a variety of configurations that are designed around the different forms of media and different connection options. Optical media libraries range from the single cartridge stand-alone model through models capable of holding 638 optical cartridges and twelve disk drives. Optical media libraries may be directly connected to the system for best functionality and performance, or may be connected through a LAN to allow independent access by PCs or other systems. Verify which adapter is appropriate for your model system and device interface.

Table 2. Currently supported optical storage devices

Model	Drive Type	Connection	Cartridge Capacity	Number of Drives
3431-705	Multi-Function	LAN	1	1
3995-A23	Multi-Function	LAN	16	1
3995-022	WORM	LAN	32	2
3995-023	Multi-Function	LAN	32	2
3995-122	WORM	LAN	144	4
3995-123	Multi-Function	LAN	144	4
3995-C20	Multi-Function	LAN	20	1 or 2
3995-C22	Multi-Function	LAN	52	2
3995-C24	Multi-Function	LAN	104	2 or 4
3995-C26	Multi-Function	LAN	156	4 or 6
3995-C28	Multi-Function	LAN	258	4 or 6
3995-C40	Multi-Function	Direct	20	1 or 2
3995-C42	Multi-Function	Direct	52	2
3995-C44	Multi-Function	Direct	104	2 or 4
3995-C46	Multi-Function	Direct	156	4 or 6
3995-C48	Multi-Function	Direct	258	4 or 6

3996-032	Multi-Function	Direct	32	2
3996-080	Multi-Function	Direct	72 or 80	2 or 4
3996-174	Multi-Function	Direct	166 or 174	2 or 4
399F-100	Multi-Function	Direct	24-80	1-4
399F-200	Multi-Function	Direct	104-638	2-12

2.2 CLASSIFICATION OF OPTICAL STORAGE DEVICES

Optical storage devices save data as patterns of dots that can be read using light. A laser beam is the usual light source.

The data on the storage medium is read by bouncing the laser beam off the surface of the medium. If the beam hits a dot it is reflected back differently to how it would be if there was no dot. This difference can be detected, so the data can be read.

Dots can be created using the laser beam (for media that is writable such as CD-Rs). The beam is used in a high-power mode to actually mark the surface of the medium, making a dot. This process is known as 'burning' data onto a disc.

2.2.1 Read-Only Optical Discs

Read-only optical discs have data written onto them when they are manufactured. This data cannot be changed.

Read-only optical media Optical storage media that cannot be written by the user but that carry data imprinted during manufacture, usually by pressing from a master disk.

CD-ROM

Compact Disc - Read-Only Memory (CD-ROM) discs can hold around 800MB of data. The data cannot be altered (non-volatile), so cannot be accidently deleted. CD-ROMs are random-access devices.

CD-ROMs are used to distribute all sorts of data: software (e.g. office applications or games), music, electronic books.

CD-ROM, abbreviation of compact disc read-only memory, type of computer memory in the form of a compact disc that is read by optical means. A CD-ROM drive uses a low-power laser beam to read digitized (binary) data that has been encoded in the form of tiny pits on an optical disk. The drive then feeds the data to a computer for processing.

The standard compact disc was introduced in 1982 for digital audio reproduction. But, because any type of information can be represented digitally, the standard CD was adapted by the computer industry, beginning in the mid-1980s, as a low-cost storage-and-distribution medium for large computer programs, graphics, and databases. With a storage capacity of 680 megabytes, the CD-ROM found rapid commercial acceptance as an alternative to so-called floppy disks (with a maximum capacity of 1.4 megabytes).

Unlike conventional magnetic storage technologies (e.g., tapes, floppy disks, and hard disks), CDs and CD-ROMs are not recordable—hence the tag "read only." This limitation spurred the development of various recordable magnetic-optical hybrid storage devices; but they generally failed to penetrate beyond the publishing world, where large multimedia files are regularly exchanged, because of incompatibility with standard CD and CD-ROM players. In the early 1990s a new type of CD became available: CD-Recordable, or CD-R. These discs differ from regular CDs in having a light-sensitive organic dye layer which can be "burned" to produce a chemical "dark" spot, analogous to an ordinary CD's pits, that can be read by existing CD and CD-ROM players. Such CDs are also known as WORM discs, for "Write Once Read Many." A rewritable version based on excitable crystals and known as CD-RW was introduced in the mid-1990s. Because both CD-R and CD-RW recorders originally required a computer to operate, they had limited acceptance outside of use as computer software and data backup devices.

To handle the proliferation of ever-larger multimedia files (audio, graphic, and video) in computer games, educational software, and electronic encyclopaedias—as well as high-definition movies for television entertainment systems—an expanded storage medium, the digital videodisc (DVD), was introduced in 1995, and a storage medium with even more capacity, Blu-ray, was introduced in 2002. However, with the increased storage capacity of computers and the easy distribution of large files over the Internet, the use of CD-ROMs declined in the 21st century.

DVD-ROM

Digital Versatile Disc - Read-Only Memory (DVD-ROM) discs can hold around 4.7GB of data (a dual-layer DVD can hold twice that). DVD-ROMs are random-access devices.

DVD-ROMs are used in the same way as CD-ROMs (see above) but, since they can hold more data, they are also used to store high-quality video.

Digital versatile disc-read only memory (DVD-ROM) is a read-only digital versatile disc (DVD) commonly used for storing large software applications. It is similar to a compact disk-read only memory (CD-ROM) but has a larger capacity. A DVD-ROM stores around 4.38 GB of data. A CD-ROM usually stores 650 MB of data.

A DVD-ROM permanently stores data files which cannot be changed, written over or erased. A personal computer (PC) with a DVD-ROM or a DVD-RAM drive is designed to read a DVD-ROM disc. Generally a DVD-ROM disc is not equipped to be used with a DVD drive connected to a home theater system or television. But many DVD-ROM drives can generally read a DVD movie disc.

A DVD-ROM is one of the various types of DVDs. A blank DVD is generally a DVD-R or DVD+R, which has a read-write format. The +R or -R references the format standards and is a rewritable or recordable DVD.

Compared to a CD-ROM, a DVD-ROM has the same 5 inch diameter and 1.2 millimeter (mm) thickness. But because a DVD-ROM uses a shorter wavelength laser with tighter compacted pits, the disc capacity is increased. In fact, the smallest DVD-ROM can store approximately 7 times more data than a CD-ROM.

The DVD-ROM was first introduced in 1996 by the DVD Forum, a group of ten international companies using and developing DVD and HD DVD formats for media, software and hardware. The DVD Forum consists of the founding companies plus over 220 other members. Japan produced the first DVD-ROMs in November 1996. By March 1997 it was introduced in the United States. The DVD Forum also releases all DVD specifications published in the DVD books by titles such as DVD-ROM Book or DVD-R Book.

A typical DVD-ROM can hold up to 17 GB/s of data if both sides of the disc are writable.

The DVD-ROM is comprised of two 0.6 millimeters (mm) acrylic layers bonded together. The double-sided disc consists of two recordable grooved sides. With two layers, a DVD's laser beam only has to go through 0.06 mm to reach the recording layer. Having a thin layer allows the lens to focus the beam to a smaller spot size, which in turn writes smaller pits for more data. The data is encoded in the form of spiral pits that are merely nanometers apart. The spiral path begins at the center of the disc and coils numerous times until it reaches the outer edge. With a double-layered disc the path continues to the second layer. A double-sided disc needs to be manually turned over and the path resumes in the center.

2.2.2 High Capacity Optical Discs

Blu-Ray

Blu-Ray disks are a recent replacement for DVDs. A Blu-Ray disc can hold 25 - 50GB of data (a dual-layer Blu-Ray disc can hold twice that). Blu-Ray discs are random-access devices.

Blu-Ray discs are used in the same way as DVD-ROMs (see above) but, since they can hold more data, they are also used to store very high-quality, high-definition (HD) video.

The 'Blu' part of Blu-Ray refers to the fact that the laser used to read the disc uses blue light instead of red light. Blue light has a shorter wave-length than red light (used with CDs and DVDs).

Using a blue laser allows more data to be placed closer together on a Blu-Ray disc, than on a DVD or CD, so Blu-Ray has a much higher storage capacity than these older discs.

A Blu-ray disk (BD) is a high-capacity optical disk medium developed for recording, rewriting and playing back high definition video. It can store large amounts of data and was designed to supersede the DVD.

Blu-ray was jointly developed by a group of personal computer and consumer electronics companies called the Blu-ray Disc Association. Blu-ray disks support higher resolutions and more advanced video and audio formats compared to DVDs.

Blu-ray technology gets its name from the blue-violet laser that is used to read Blu-ray disks. Compared to a DVD's red laser, a blue laser permits more information to be stored at a greater density. For example, while a DVD can store 15 GB per layer, a Blu-ray disk can store 25 GB per layer, and dual-layer disks can hold up to 50 GB.

Compared to a DVD, Blu-ray also provides much higher resolution; while a DVD with standard definition can provide definition of 720x480 pixels, Blu-ray high definition has 1920X1080 pixel resolution.

HD DVD

High-density DVD (HD-DVD) discs can hold around 15GB of data (a dual-layer HD-DVD can hold twice that). HD-DVDs are random-access devices.

HD-DVD discs are used in the same way as DVD-ROMs (see above) but, since they can hold more data, they are also used to store very high-quality, high-definition (HD) video.

The HD-DVD format was launched at the same time as Blu-Ray. For about a year they competed to be the 'next DVD'. For various reasons, Blu-Ray won the fight, and the HD-DVD format has been abandoned.

An HD-DVD uses the same principles -- it contains a bumpy layer that reflects light from a laser to a sensor, creating a digital signal. HD-DVDs are even exactly the same size as DVDs (120 millimeters in diameter and 1.2 millimeters thick). But three important differences allow them to hold quite a bit more information than DVDs:

- They use 405-nanometer blue-violet lasers rather than 650-nanometer red lasers.
- The pits are smaller and the tracks are closer together.
- They use more efficient compression to cut down the size of the files they store.

The color of the laser may seem like a trivial change to make, but the shorter wavelength of the blue-violet laser is what allows HD-DVD's pits to be smaller and arranged closer together. In other words, it allows the disc to have a much narrower track pitch. Regular DVDs have a track pitch of 0.74 micrometers, and HD-DVDs have a track pitch of 0.40 micrometers. You can imagine this as the difference between writing with a magic marker and a fine-tipped pen.

The other big difference between DVDs and HD-DVDs involves how the information on the disc is compressed. Most DVDs use MPEG-2 compression. HD-DVDs can use MPEG-2, but they typically use the more efficient MPEG-4, which allows higher video quality with a smaller file size. HD-DVDs can also use VC-1 (or Windows Media) compression.

Finally, because of general improvements in the technology, an HD-DVD player can read information from the disc and deliver it to the TV about three times as fast as a DVD player can. It can also send the signal to an HDTV digitally using a High Definition Multimedia Interface (HDMI), preventing the quality loss that conversion to analog causes.

One of the first questions people ask about HD-DVD (besides “Is it better than Blu-ray?”) is whether their old DVDs are about to become obsolete. Let’s take a look at what is likely to happen with players and discs as people upgrade.

2.2.3 Recordable Optical Discs

Recordable optical discs can have data written onto them (‘burnt’) by a computer user using a special disc drive (a disc ‘burner’).

Recordable optical disc media contains an organic dye layer whose transparency can be altered either to absorb a laser beam or to allow the beam to pass through to a reflective layer behind the dye. The nature of this organic dye is such that when the internal energies of its molecules reach a particular threshold, an irreversible chemical reaction occurs, and the dye layer loses its transparency. This property allows a high-energy beam to “write” data by burning “pits,” in the form of dark marks, to the disc during recording. A low powered laser reads the data by either passing through the transparent dye layer (without causing any molecular change) to the reflective layer or by being absorbed by the nontransparent marks in the dye. Due to the organic nature of the dye, degradation and breakdown of the transparent portion of dye layer will occur over a long period of time as a natural process. This process, which has its roots in chemical kinetics, can take several years in normal environment conditions. Higher temperatures and humidity will accelerate this process by increasing the thermal and kinetic energies of the dye molecules. It is well known that temperature and humidity are among the most important factors affecting the life expectancy of optical discs. Yet, there is another important factor that has not been so well investigated. Light exposure can increase the rate of dye degradation precisely because the organic dye used in recordable media is light sensitive. This study also addresses this issue.

CD-R and DVD-R

CD-Recordable (CD-R) and DVD-recordable (DVD-R) discs can have data burnt onto them, but not erased. You can keep adding data until the disc is full, but you cannot remove any data or re-use a full disc.

DVD-R and CD-R are two devices used to store data. While DVD-R stands for Digital Versatile Discs-Recordable, CD-R stands for Compact Disc-Recordable. The basic difference between the two lies in the amount of data they can store. While a CD is able to store only 700 MB data, a DVD-R can hold up to 4.7 GB of data. Both DVD-R and CD-R can be used to store audio and video. While the capacity of 4.7 GB of a DVD-R translates into 120 minutes of video, 700 MB capacity of a CR-R means you can store about 80 minutes of audio data in them. For backing up your data, DVD-R is preferred because of its capacity and for simple economics also.

DVD's and CD's come in blank forms allowing a user to fill data on them. Talking of differences, the data in a DVD is advanced in quality and sharpness and video files in DVD's are of HD quality, and if you wish to store films, it is better to use a DVD-R instead of a CD-R as not only is the video quality better, more space means you can store up to 5 films in a DVD-R, while a CD-R can hold only one film.

To burn a DVD-R, you require a DVD burner which is software that allows burning of audio or video files onto your DVD-R. This DVD burner also has the capacity to burn a CD. The advantage between the two is visible when you are trying to watch a DVD or CD on your DVD player. It is very much possible as a DVD player can read both CD and DVD but a CD player cannot read a DVD.

Earlier, there was a huge price difference between DVD-R and CD-R. But with advancement of technology, the prices of DVD-R and CD-R have almost become similar. CD's are still somewhat cheaper than DVD's and this is the only advantage they hold over DVD's.

Today there are available both DVD-RW and CD-RW which are rewritable which means you can record and then erase them to use them again for storage. However, even DVD-R and CD-R have limited life and you cannot keep on writing on them.

CD-RW and DVD-RW

CD-ReWritable (CD-RW) and DVD-ReWritable (DVD-RW) discs, unlike CD-Rs and DVD-Rs, can have data burnt onto them and also erased so that the discs can be re-used.

When CD-Rs and DVD-Rs are burnt, the laser makes permanent marks on the silver-coloured metal layer. This is why these discs cannot be erased.

When CD-RWs and DVD-RWs are burnt the laser makes marks on the metal layer, but in a way that can be undone. So these discs can be erased.

DVD-RAM

DVD-Random Access Memory (DVD-RAM) discs are a type of re-writable DVD. They often come in a floppy-disc style case (to protect the disc).

DVD-RAM discs have a similar capacity to a normal DVD, holding 4.7GB of data. DVD-RAM discs are random-access devices.

DVD-RAM discs are used in many camcorders (video recording cameras).

The discs are much higher quality than normal DVD-RWs and can reliably store data for up to 30 years. This means that they are often used for video and data back-up and archiving.

DVD-RAM is a portable, removable and rewritable optical disc format. Unlike the standard DVD medium, DVD-RAM can be written, erased and/ or overwritten (with a maximum overwrite capacity of 100,000). The format uses phase change recording, in which alternating laser intensity changes the discs' record layers to and from various physical states, and is only compatible with devices that are specifically manufactured to support it.

The DVD-RAM format was first presented by the DVD Forum in 1996 and the first DVD-RAM drive was introduced by Panasonic two years later. Since its origin, the medium has been available as single- and double-sided discs. DVD-RAM drives are typically able to also read DVD-ROM, standard DVDs and all different forms of compact disc. There now exist at least a dozen different specifications for the format, with a wide range of recording speeds and storage capacities.

2.3 OPTICAL MEDIA FORMATS STRATEGY

The media format is the file system architecture that exists on the media to manage file, directory, and volume information.

Writable optical media (such as WORM, erasable, DVD-RAM) is initialized on IBM i using the Initialize Optical (INZOPT) command. WORM media must use the High Performance Optical File System (HPOFS) format. DVD-RAM media must use Universal Disk Format (UDF). Erasable media can use either HPOFS or UDF depending on the requirements of the user. You can specify the format by using the MEDFMT keyword on the Initialize the optical volume (INZOPT) command. The following topics provide information about the different media formats and a comparison so you can select the media format that best meets your requirements.

- **ISO 9660** This industry standard media format was originally designed to specify the volume and file structures of compact-disk read-only memory (CD-ROM) optical disks, and is a read-only media format.
- **High performance optical file system** High Performance Optical File System (HPOFS) is an IBM-developed media format architecture available to use when initializing optical media.
- **Universal Disk Format** Universal Disk Format (UDF) is the Optical Storage Technology Association (OSTA) supported group of ISO/ IEC 13346.
- **CL command support for media formats** Use these commands to save and restore data. And, read about the restrictions for the ISO 9660, High Performance Optical File System (HPOFS), and Universal Disk Format (UDF) media.

2.3.1 Configuring Optical Devices

Configure your CD-ROM, DVD-ROM, and DVD-RAM devices, and optical medial libraries. The systems come with a rack-mounted CD-ROM or DVD-ROM drive.

As an option, you can order a DVD-RAM drive as a feature to replace your internal drive or to have it in addition to your internal drive. All optical drives are multi-user devices that multiple users can access concurrently.

Directly attached optical media libraries are attached to the system through the small computer system interface (SCSI). On IBM® i, multiple users can access data on optical media libraries concurrently. These interfaces are available to access the data on the CD and DVD devices and optical media libraries:

- Save and restore interface
- Hierarchical file system (HFS)
- Application programming interface (API)
- Integrated file system interfaces
- Optical commands and utility displays

2.3.2 Connectivity of Library Devices

In addition to optical libraries, you can now attach some non-optical library devices to the system. For more information about which devices are supported and system configuration requirements.

- **Configuring your optical drive.** Depending on the model of your system, you can position the CD-ROM or DVD drive either horizontally or vertically in the system.
- **Configuring directly attached optical media libraries.** To create a device description for an optical media library device, use the Create Device Description (Media Library) (CRTDEVMLB) command.
- **Labeling optical cartridges.** Each optical disk cartridge contains two sides. Each side corresponds to an optical volume. There are two techniques to associate a volume ID label with the correct side of the optical cartridge. This is important to know when you set the write-protect switch.
- **Getting started with optical cartridges and volumes.** Read about the optical support functions and familiarize yourself with using optical cartridges and volumes.

2.3.3 Using Optical Devices

You can display the primary menu for optical support by entering GO OPTICAL on the IBM® i command line. System administrators and programmers can access most optical commands through this menu. It is also convenient to enter many of the optical commands directly on the command line.

These commands offer the following functions:

- Display optical volumes in a directly attached or LAN-attached optical media library device (MLD), CD device, DVD device, or virtual optical device.
- Display files and directories that are contained in any directory in any optical volume.
- Display the file attributes of any optical file.
- Import or export media in a directly attached optical media library, CD-ROM device, DVD device, or virtual optical device.

- Make backup copies of volumes, directories, or files that are contained in directly attached optical devices or virtual optical devices.
- Initialize a volume that is contained in a DVD-RAM drive, a directly attached optical media library, or a virtual optical device.
- Work with devices that represent optical media libraries, optical systems, CD drives, DVD drives, and virtual optical drives.
- Add, remove, or change the status of any LAN-attached optical system.
- Duplicate one optical volume to another.
- Copy files and directories from one optical volume to another.
- Check a volume for damaged directories and files.

When you enter GO CMDOPT on the command line, a complete list of optical commands appears. Many of these commands are accessible through the previous GO OPTICAL menu.

2.3.4 Using Optical Volumes

These topics describe the Work with Optical Volumes options on the Optical Support Utilities main menu.

These options are organized hierarchically, with volumes as the highest in the order and files as the lowest in the order. You can use the appropriate "Work with..." command to access these panels directly without having to go through the Optical Support Utilities main menu. Each display presents the selected information and the options that are available. Some options might not apply to all optical devices or volumes.

The primary menu for working with optical volumes is the Work with Optical Volumes display. There are several variations of the display to accommodate alternative formats and extended attribute information.

You can select the Work with Optical Volumes display by choosing Work with optical volumes on the Optical Support Utilities menu. You can also run the Work with Optical Volumes (WRKOPTVOL) command on the command line.

The Work with Optical Volumes (WRKOPTVOL) command applies to the following volumes:

- Displaying optical volumes. When the Work with Optical Volumes display first appears, it includes a list of all volumes in all CD-

ROM devices, DVD devices, RMS devices, optical media libraries, and LAN-attached devices.

- Initializing optical volumes. You must initialize the writable optical media before the system can create directories and files.
- Renaming optical volumes. You can rename an optical volume without losing the information about the volume.
- Adding optical disk cartridges. Use this procedure to add an optical disk cartridge to an optical volume.
- Copying optical volume data. Optical files can be copied from one or more volumes or directories to other volumes or directories.
- Output file support structures. There are three possible record formats that are created by the Display Optical (DSPOPT) command when output is directed to either an output file or user space.
- Changing optical volume attributes. You can change the optical volume attributes with the Change Optical Volume (CHGOPTVOL) command.
- Displaying and printing optical volume attributes. These topics provide information about viewing and printing the attributes of a volume.
- Duplicating optical volumes. An efficient method to create a backup of an optical volume is to use the Duplicate Optical (DUPOPT) command. This command copies sectors to create a volume that is identical to the source except for the volume identifier and creation date and time.
- Viewing directory and file information. There are two commands that you can use to view directory and file information through the optical support panels: the Work with Object Links (WRKLNK) command and the Work with Optical Directories (WRKOPTDIR) command.
- Removing and deleting optical volumes. You can remove optical volumes from an optical disk cartridge and then delete removed volumes from the optical index database.
- Checking optical volume. You can use the Check Optical Volume (CHKOPTVOL) command to validate the integrity of the directories and files on a volume.
- Changing optical environment parameters. The Change Optical Attributes (CHGOPTA) command can be used to change specific optical configuration parameters that affect all jobs using the optical file system.

- Settings write protection. The write-protect function prevents writing to disk. A write-protect window shows when write protection is on or off.
- Creating a master CD-ROM. The links provided will serve as a reference for instructions about CD premastering.

2.3.5 Optical Volume Backup

Use the information provided to define your backup strategy, learn about backup options, and learn backup commands.

- **Defining a backup strategy**. There is no one perfect backup strategy that meets everyone's needs. Therefore, it is important to define your backup requirements before you decide on a backup strategy. Use the questions in this topic to help you determine your backup requirements.
- **Using the Duplicate Optical (DUPOPT) command**. The Duplicate Optical (DUPOPT) command can be used to create a duplicate optical volume.
- **Copy Optical (CPYOPT) command**. Use the Copy Optical (CPYOPT) command to copy optical files and directories between optical volumes.
- **Save and restore commands**. The save (SAV) command can be used to create a backup of an optical volume image. The volume image is restored using the restore (RST) command.

2.3.6 Managing Performance in Optical Media Libraries

Several factors can affect the optical performance of both LAN-attached and directly attached optical media libraries.

Volume mounting and dismounting. Volume mounting and dismounting are important factors that affect optical performance. It takes approximately 8 to 15 seconds to remove a volume, store it in a slot, retrieve a new volume, and mount it. If you can minimize the number of volume mounts and dismounts that your application requires, optical performance will improve.

Drive contention. Performance can be severely affected by drive contention. The following conditions increase drive contention and should be avoided:

- Only one drive is available for use by applications libraries.
- Many optical processes are running that attempt to use different optical volumes at the same time.

Number of directories and files. Performance can be affected by having too few directories with too many files. Directories group related information to provide a means of quicker access. Typically, you get better performance from more directories with fewer files. Although there is no enforced limit on how many files there can be in a directory, you probably should not have more than 6000 for performance reasons.

File size. The size of a file has a direct effect on the amount of time it takes to read, write, or copy the file. In general, the larger the file, the longer the operation can be expected to take.

Add optical cartridge performance. Specifying *NO for the Rebuild Directory Index can improve the performance of Add Optical Cartridge (ADDOPTCTG) by deferring the build of the optical directory index until a later time.

2.3.7 Managing Optical Security and Auditing

The level of security available depends on the optical media format of the volume. You can use an authorization list to secure all optical volumes. This includes all volumes in CD-ROM, DVD, LAN-attached, directly-attached, and virtual optical devices. Universal Disk Format (UDF) volumes provide directory- and file-level security in addition to authorization list security. Optical support provides ways to prevent unauthorized access and processing of data that is stored on optical volumes. Optical support does this by verifying a requester's rights to specific optical volumes before attempting the following requests:

- Open file or directory
- Create directory
- Delete file or directory
- Rename file
- Initialize or rename volume
- Remove cartridge
- Change or retrieve attributes
- Copy
- Backup or convert backup

- Save or release held files
- Read sectors
- Save optical volume storage
- Restore optical volume storage
- Check optical volume for damaged files

Along with security for optical volumes, directories, and files, auditing of access to optical objects is also available.

2.3.8 Reclaiming the Optical Index Database

A system-level index, called the optical index database, keeps track of all optical volumes and directories known to the system.

The optical index database includes the optical volume index (QAMOVAR) and the optical directory index (QAMOPVR) physical files. You can use the Reclaim Optical (RCLOPT) command to re-create the optical index database if it is ever damaged or destroyed or whenever volumes that you know are in an optical media library, CD-ROM, or DVD device are reported as not found. To run the RCLOPT command, either select option 2 (Reclaim optical index) on the Optical Backup/Recovery display or enter the RCLOPT command. Doing either causes the Reclaim Optical (RCLOPT) display to appear.

- **Reclaiming the optical index for a stand-alone optical device**. The optical index database re-creates entries for CD-ROM and DVD devices each time the device is varied on with media in the device.
- **Reclaiming types**. There are three possible types to select: *SYNC, *UPDATE, and *RESET.
- **Optical index information**. Optical index information regarding which volumes are in a particular optical media library and which directories are on each volume is kept at different levels within the system.
- **Choosing the reclaim type to use**. Decide which reclaim type to use and when the different options should be used.

2.3.9 Saving and Restoring using Optical Media

Optical media is a cost effective long-term storage solution. You can save and restore your data using optical media in multiple ways. You can perform saves using BRMS, save operations and restore commands, and the Load Run command.

The IBM® i Save and Restore commands support directly attached optical media library devices, CD-ROM, DVD-ROM, DVD-RAM stand-alone devices, and virtual optical devices. The best use of optical storage devices is for disaster recovery protection. The extraordinary long shelf life of optical media is well suited for the long-term storage of critical data. You can provide extra protection by using permanent WORM media because you cannot alter data on the media. Tape devices may provide the optimal day-to-day backup mechanism. This depends on the amount of data that you want backed up, and the amount of system time available for backup. CD-ROM and DVD-RAM media are also well suited for software distribution. The save/restore command interface can be used as a part of installation procedures for programs, data, and program fixes. CD-ROM and DVD-RAM stand-alone drive optical devices also support the Load Run (LODRUN) command.

Backup, Recovery and Media Services (BRMS) is a licensed program that helps you create a disciplined approach to manage your backups. Optical media is supported by BRMS. Refer to the BRMS topic for more details.

- **Concepts for saving and restoring optical media**. Optical devices support many of the most widely used IBM i Save and Restore commands.
- **Saving optical files to optical storage**. You can save optical files to DVD-RAM, UDF, and HPOFS media.
- **Operating guidelines by optical device type**. There are operational guidelines by device type for optical library data servers and CD-ROM, DVD-ROM, and DVD-RAM stand-alone optical drive devices.

2.4 FUTURE TENDENCY OF OPTICAL STORAGE TECHNOLOGY

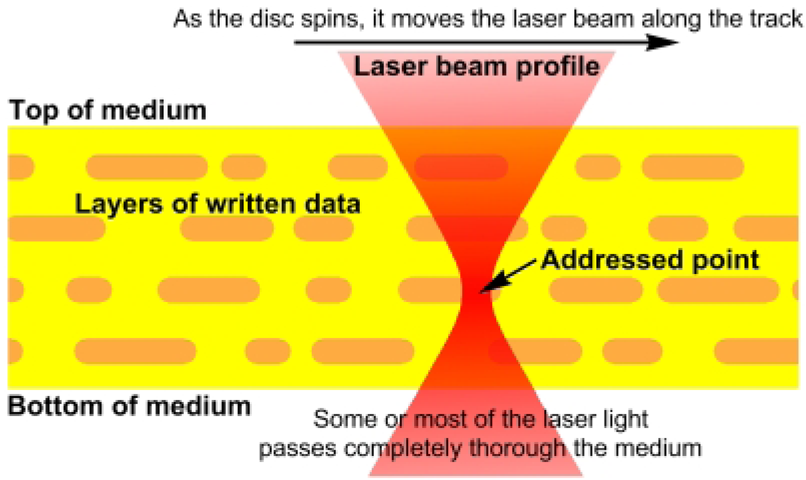
The origins of the field date back to the 1950s, when Hirshberg developed the photo chromic spiropyrans and suggested their use in data storage. In the 1970s, Barachevskii demonstrated that this photochromism could be produced by two-photon excitation, and finally at the end of the 1980s Peter T. Rentzepis showed that this could lead to three-dimensional data storage. This proof-of-concept system stimulated a great deal of research and development, and in the following decades many academic and commercial groups have worked on 3D optical data storage products and technologies. Most of the developed systems are based to some extent on the original ideas of Rentzepis. A wide range of physical phenomena for data reading and recording have been investigated, large numbers of chemical systems for the medium have been developed and evaluated, and extensive work has been carried out in solving the problems associated with the optical systems required for the reading and recording of data. Currently, several groups remain working on solutions with various levels of development and interest in commercialization.

One of the reasons that computers have become increasingly important in daily life is because they offer unprecedented access to massive amounts of information. The decreasing cost of storing data and the increasing storage capacities of ever smaller devices have been key enablers of this revolution. Current storage needs are being met because improvements in conventional technologies such as magnetic hard disk drives, optical disks, and semiconductor memories have been able to keep pace with the demand for greater and faster storage.

However, there is strong evidence that these surface-storage technologies are approaching fundamental limits that may be difficult to overcome, as ever-smaller bits become less thermally stable and harder to access. Exactly when this limit will be reached remains an open question: some experts predict these barriers will be encountered in a few years, while others believe that conventional technologies can continue to improve for at least five more years. In either case, one or more successors to current data storage technologies will be needed in the near future.

An intriguing approach for next generation data-storage is to use light to store information throughout the three-dimensional volume of a material. By distributing data within the volume of the recording medium, it should be possible to achieve far greater storage densities than current technologies can offer.

For instance, the surface storage density accessible with focused beams of light is roughly 1/ (2 Wave length). With green light of roughly 0.5 micron wavelength, this should lead to 4 bits/sq. micron or more than 4 Gigabytes (GB) on each side of a 120mm diameter, 1mm thick disk. But by storing data throughout the volume at a density of 1/ (3Wave length), the capacity of the same disk could be increased 2000 fold, to 8 Terabytes (TB).



Schematic representation of a cross-section through a 3D optical storage disc (yellow) along a data track (orange marks). Four data layers are seen, with the laser currently addressing the third from the top. The laser passes through the first two layers and only interacts with the third, since here the light is at a high intensity.

Current optical data storage media, such as the CD and DVD store data as a series of reflective marks on an internal surface of a disc. In order to increase storage capacity, it is possible for discs to hold two or even more of these data layers, but their number is severely limited since the addressing laser interacts with every layer that it passes through on the way to and from the addressed layer. These interactions cause noise that limits the technology to perhaps ~10 layers. 3D optical data storage methods circumvent this issue by using addressing methods where only the specifically addressed voxel interacts substantially with the addressing light. This necessarily involves nonlinear data reading and writing methods, in particular nonlinear optics. 3D optical data storage is related to (and competes with) holographic data storage, but operates on different principles.

As an example, a prototypical 3D optical data storage system may use a disk that looks much like a transparent DVD. The disc contains many layers of information, each at a different depth in the media and each consisting of a DVD-like spiral track. In order to record information on the disc a laser is brought to a focus at a particular depth in the media that corresponds to a particular information layer. When the laser is turned on it causes a photochemical change in the media. As the disc spins and the read/write head moves along a radius, the layer is written just as a DVD-R is written. The depth of the focus may then be changed and another entirely different layer of information written. The distance between layers may be 5 to 100 micrometers, allowing >100 layers of information to be stored on a single disc.

In order to read the data back, a similar procedure is used except this time instead of causing a photochemical change in the media the laser causes fluorescence. This is achieved e.g. by using a lower laser power or a different laser wavelength. The intensity or wavelength of the fluorescence is different depending on whether the media has been written at that point, and so by measuring the emitted light the data is read.

2.4.1 Processes for Writing Data

Data recording in a 3D optical storage medium requires that a change take place in the medium upon excitation. This change is generally a photochemical reaction of some sort, although other possibilities exist. Chemical reactions that have been investigated include photoisomerizations, photodecompositions and photo bleaching, and polymerization initiation. Most investigated have been photochromic compounds, which include azobenzenes, spiropyrans, stilbenes, fulgides and diarylethenes. If the photochemical change is reversible, then rewritable data storage may be achieved, at least in principle. Also, multilevel recording, where data is written in 'grayscale' rather than as 'on' and 'off' signals, is technically feasible.

Although there are many nonlinear optical phenomena, only multiphoton absorption is capable of injecting into the media the significant energy required to electronically excite molecular species and cause chemical reactions. Two-photon absorption is the strongest multiphoton absorbance by far, but still it is a very weak phenomenon, leading to low media sensitivity. Therefore, much research has been directed at providing chromophores with high two-photon absorption cross-sections.

Two-photon absorption

Writing by 2-photon absorption can be achieved by focusing the writing laser on the point where the photochemical writing process is required. The wavelength of the writing laser is chosen such that it is not linearly absorbed by the medium, and therefore it does not interact with the medium except at the focal point. At the focal point 2-photon absorption becomes significant, because it is a nonlinear process dependant on the square of the laser fluence.

Writing by 2-photon absorption can also be achieved by the action of two lasers in coincidence. This method is typically used to achieve the parallel writing of information at once. One laser passes through the media, defining a line or plane. The second laser is then directed at the points on that line or plane that writing is desired. The coincidence of the lasers at these points excited 2-photon absorption, leading to writing photochemistry.

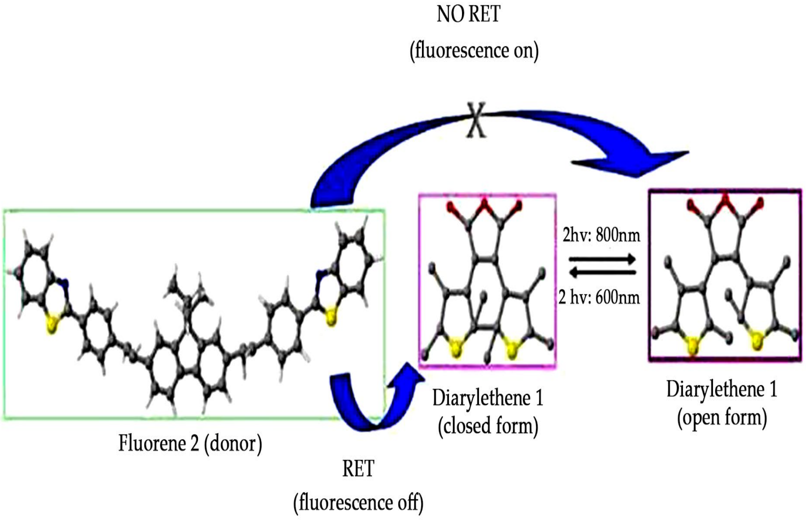
Another approach to improving media sensitivity has been to employ resonant two-photon absorption. Nonresonant two-photon absorption (as is generally used) is weak since in order for excitation to take place, the two exciting photons must arrive at the chromophore at almost exactly the same time. This is because the chromophore is unable to interact with a single photon alone. However, if the chromophore has an energy level corresponding to the (weak) absorption of one photon then this may be used as a stepping stone, allowing more freedom in the arrival time of photons and therefore a much higher sensitivity. However, this one-photon absorbance is a linear process, and therefore risks compromising the 3D resolution of the system.

Two photon absorption (TPA) is the simultaneous absorption of two photons of identical or different frequencies in order to excite a molecule from its ground state to an excited state. The first TPA process was observed in doped europium salts

Two-photon absorption can be measured by several techniques. Two of them are two-photon excited fluorescence (TPEF) and nonlinear transmission (NLT). Pulsed lasers are most often used because TPA is a third-order nonlinear optical process, and therefore is most efficient at very high intensities.

In non resonant TPA two photons combine to bridge an energy gap larger than the energies of each photon individually. If there were an intermediate state in the gap, this could happen via two separate one-photon transitions in a process described as "resonant TPA", "sequential TPA", or "1+1 absorption". In non resonant TPA the transition occurs without the presence of the intermediate state. The "nonlinear" in the description of this process means that the strength of the interaction increases faster than linearly with the electric field of the light. In fact, under ideal conditions the rate of TPA is proportional to the square of the field intensity. This dependence can be derived quantum mechanically, but is intuitively obvious when one considers that it requires two photons to coincide in time and space. This requirement for high light intensity means that lasers are required to study TPA phenomena.

Further, in order to understand the TPA spectrum, monochromatic light is also desired in order to measure the TPA cross section at different wavelengths. Hence, tunable pulsed lasers (such as frequency-doubled Nd: YAG-pumped OPOs and OPAs) are the choice of excitation.



Description: A two-photon 3D optical data storage system consisting of a bichromophoric mixture of diarylethene and fluorene derivative as the storage medium is demonstrated here. Binary information bits were recorded throughout all three dimensions of the storage medium by two-photon localized excitation on the diarylethene molecules, transforming the closed form of diarylethene into the open form. The readout method is based on the modulation of the two-photon fluorescence emission of fluorene by the closed form of diarylethene

Micro fabrication

One of the most distinguishing features of TPA is that the rate of absorption of light by a molecule depends on the square of the light's intensity. This is different than OPA, where the rate of absorption is linear with respect to input intensity. As a result of this dependence, if material is cut with a high power laser beam, the rate of material removal decreases very sharply from the center of the beam to its periphery. Because of this, the "pit" created is sharper and better resolved than if the same size pit were created using normal absorption. In the case of two-photon polymerization, the material is polymerized only near the focal spot of the laser, where the intensity of the absorbed light is highest. This makes TPA attractive for 3D micro fabrication.

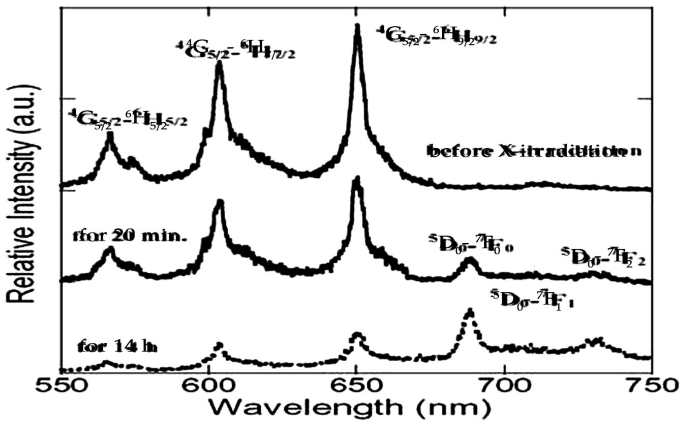
Data recording during manufacturing

Data may also be created in the manufacturing of the media, as is the case with most optical disc formats for commercial data distribution. In this case, the user can not write to the disc - it is a ROM format. Data may be written by a nonlinear optical method, but in this case the use of very high power lasers is acceptable so media sensitivity becomes less of an issue.

The fabrication of discs containing data molded or printed into their 3D structure has also been demonstrated. For example, a disc containing data in 3D may be constructed by sandwiching together a large number of wafer-thin discs, each of which is molded or printed with a single layer of information. The resulting ROM disc can then be read using a 3D reading method.

Persistent Spectral Hole Burning (PSHB): Persistent spectral hole-burning has been utilized as a means for possibly achieving high-density optical storage, which also allows the possibility of spectral multiplexing to increase data density. Persistent spectral holes are formed in inhomogeneously broadened absorption lines when a photo induced change occurs in the subset of absorbers that are in resonance with a narrowband laser beam. If the photo reacted centers do not absorb at the original wavelength, a dip in absorption or spectral 'hole' is formed that may be detected by subsequent measurement of the absorption line.

Divalent samarium (Sm2+) and trivalent europium (Eu3+) ions in glasses are of special importance for their properties of persistent spectral hole-burning (PSHB), which is promising as an extremely high-density optical memory using a wavelength region in addition of spatial two-dimensions. PSHB of the rare-earth ions with 4*f* 6 configuration is conceptually based on a single site excitation and photochemical reaction of the ions in their inhomogeneous distribution of 5D0-7F0 energies in glasses. The homogeneous line width is ~ 0.1 cm-1 at 77 K, high density data storage at a light spot (~1μm), of ~ 30 bit/spot at room temperature and ~ 1000 bit/spot at 77 K, may be achieved. It is believed that the PSHB of Sm2+ ions is a photo-ionization of Sm2+ **+** *h*ν giving Sm3+ + *e*-; the electron generated is captured in a defect site in glasses neighbouring to the photo-reacted Sm2+ and a persistency of the spectral hole with very narrow homogeneous width is eventually obtained.



However, PSHB media currently requires extremely low temperatures ranging from 1.5K to50K to be maintained in order to avoid data loss.

Microholography: where tiny holograms are used to store data. In micro holography, focused beams of light are used to record submicron-sized holograms in a photorefractive material, usually by the use of collinear beams. The writing process may use the same kinds of media that are used in other types of holographic data storage, and may use 2-photon processes to form the holograms.

Void Formation: where microscopic bubbles are introduced into a media by high intensity laser irradiation. Standard set-up consists of a laser providing amplified femtosecond pulses and an optical microscope is used for recording voids inside glasses under tight focusing conditions using an objective lens with a numerical aperture of NA = 1.35. The diameter of the focal spot was estimated as $D = 1.22\lambda/\mathrm{NA}$ at the $1/e2$-level by intensity.

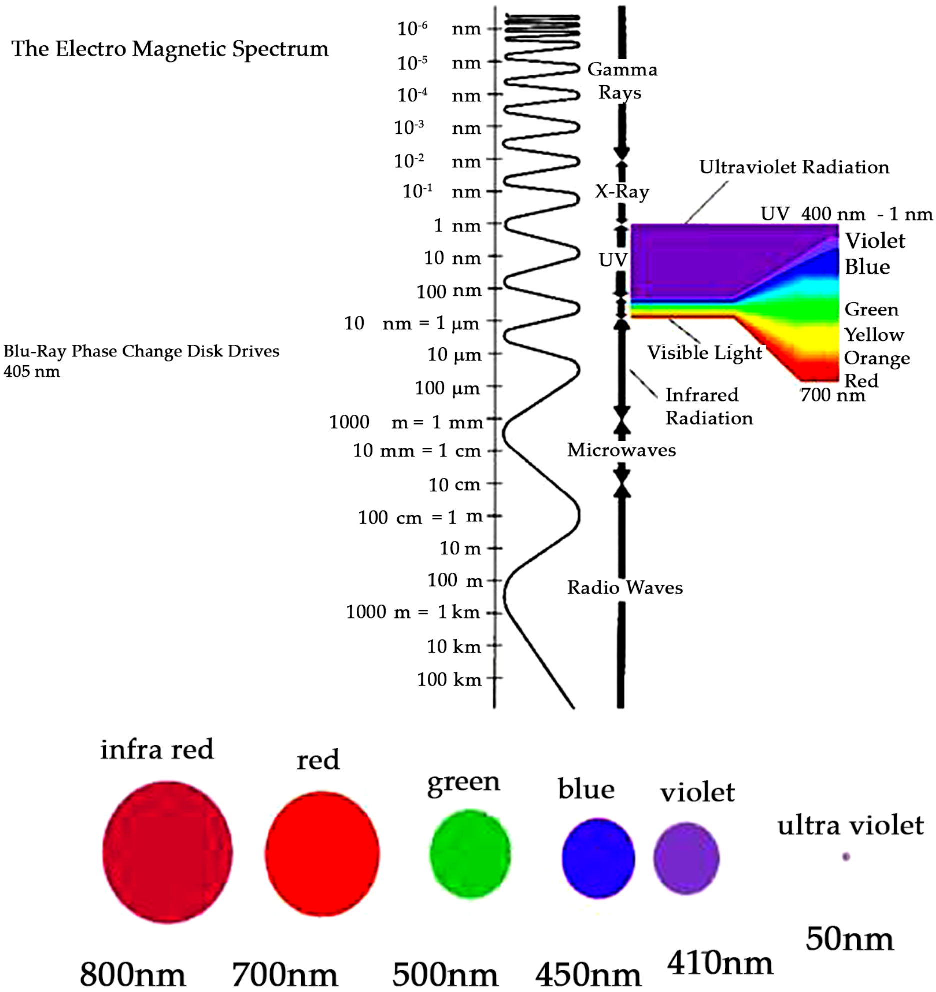
The void inside glass represents a kind of ultimate density modulation created by a laser pulse: the empty volume surrounded by a shell of densified material It is technologically important to establish the conditions of formation of such photo-modification for micro-structuring, creation of new phases inside the densified region with altered chemical properties, as well as to assess damage resistance of glasses at extremely high irradiance (> 10 TW cm–2). Voids of sub-micrometer cross-sections can be recorded inside glass by tightly focused single ultra-short pulses without crack formation.

Chromophore Poling: where the laser-induced reorientation of chromophores in the media structure leads to readable changes. Optical engineering of the photonic properties of polymer films provides efficient and convenient ways to store information within photo-sensitive materials.

Azo-dye doped or side-chain polymers have attracted tremendous attention over the years because of their photo physical properties which can be exploited in many non linear optical applications. In this context, azo-doped poly (methyl methacrylate) (PMMA) polymer film is a model system which has been widely exploited for optical switching, holographic storage or optical memories where the required order of the chromophores can be achieved not only by applying an electric field but also via optical poling. In the field of optical data storage, the crucial parameter is spatial resolution. This need has been specifically addressed by the use of femtosecond laser sources, which, because of their high peak power, are able to give rise to multiphoton processes localized within a sub-wavelength volume in the vicinity of the focal point. This has been achieved by performing orientational hole burning through two-photon absorption in films of poly (methyl methacrylate) doped with Disperse Red 1 (DR1) that can subsequently be detected through confocal differential reflexion microscopy. A more sophisticated approach has recently been proposed by the Zyss group which encodes information by an all-optical poling technique in which the angles of polarization of the two irradiating fields are varied. The resulting spatial changes in the symmetry of the quadratic susceptibility tensor lead to a modulation of the detected SHG intensity when scanning the sample with the IR beam alone thus using a nonlinear optical phenomenon for the read-out stage as well. One possible way to combine the advantages of these two distinct techniques would consist in using two-photon isomerisation in a photo-assisted poling scheme followed by a simple SHG read-out stage. However, the critical step would remain the orientation of chromophores in a small volume and over a limited time span. Therefore, we propose to start out with the even simpler method of writing optical data into a previously corona poled film by locally disorienting the polar order, now using two photon isomerisation to randomize the initial orientation of the chromophores. Again, data retrieval will be performed by monitoring SHG intensity while scanning the sample with an IR beam. In our approach, the information is being encoded into the succession of localized areas which have been disordered or not. This takes advantage of the fact that it is by far easier to induce disorder than to create order, and that the former is more irreversible. In addition, we will show that the intensity thresholds will be low enough to allow the erasing of the data by heating the sample and the rewriting of new data after repoling it.

Role of wave length

Amount of data that can be stored is dependent on the wave length of light used. Optical refraction limits the size of focused laser beam, so a spot of the order of the wave length is used represent the presence of information; therefore wave length limits the density of data storage.



mm=nanometer

2.4.2 Processes for Reading Data

The reading of data from 3D optical memories has been carried out in many different ways. While some of these rely on the nonlinearity of the light-matter interaction to obtain 3D resolution, others use methods that spatially filter the media's linear response. Reading methods include:

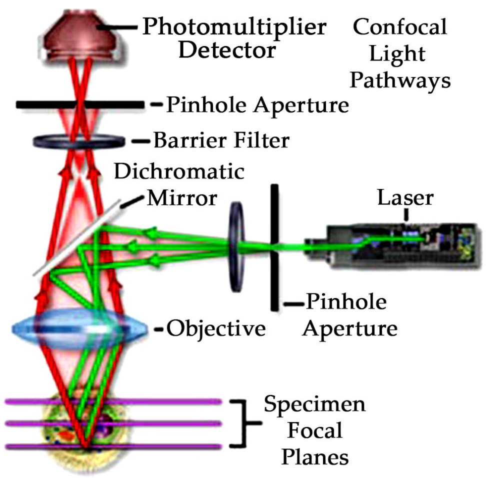
1. **TWO photon EXCITATION FLUORESCENCE.** This method is essentially two-photon microscopy. Two-photon excitation may in some cases be a viable alternative to confocal microscopy due to its deeper penetration and reduced photo toxicity.

Two-photon excitation employs a concept first described by Maria Göppert-Mayer (b. 1906) in her 1931 doctoral dissertation. The concept of two-photon excitation is based on the idea that two photons of low energy can excite a fluorophore in a quantum event, resulting in the emission of a fluorescence photon, typically at a higher energy than either of the two excitatory photons. The probability of the near-simultaneous absorption of two photons is extremely low. Therefore a high flux of excitation photons is typically required, usually a femtosecond laser.

Two-photon microscopy was pioneered by Winfried Denk in the lab of Watt W. Webb at Cornell University. He combined the idea of two-photon absorption with the use of a laser scanner. In two-photon excitation microscopy an infrared laser beam is focused through an objective lens. The Ti-sapphire laser normally used has a pulse width of approximately 100 femtosecond and a repetition rate of about 80 MHz, allowing the high photon density and flux required for two photons absorption and is tunable across a wide range of wavelengths. Two-photon technology is patented by Winfried Denk, James Strickler and Watt Webb at Cornell University. He most commonly used fluorophores (A fluorophore, is a component of a molecule which causes a molecule to be fluorescent. It is a functional group in a molecule which will absorb energy of a specific wavelength and re-emit energy at a different (but equally specific) wavelength) have excitation spectra in the 400–500 nm range, whereas the laser used to excite the fluorophores lies in the ~700–1000 nm (infrared) range. If the fluorophore absorbs two infrared photons simultaneously, it will absorb enough energy to be raised into the excited state. The fluorophore will then emit a single photon with a wavelength that depends on the type of fluorophore used (typically in the visible spectrum). Because two photons need to be absorbed to excite a fluorophore, the probability for fluorescent emission from the fluorophores increases quadratically with the excitation intensity. Therefore, much more two-photon fluorescence is generated where the laser beam is tightly focused than where it is more diffuse. Effectively, fluorescence is observed in any appreciable amount in the focal volume, resulting in a high degree of rejection of out-of-focus objects. The fluorescence from the sample is then collected by a high-sensitivity detector, such as a photomultiplier tube. This observed light intensity becomes one pixel in the eventual image; the focal point is scanned throughout a desired region of the sample to form all the pixels of the image.

2. **Confocal detection.** This method is essentially confocal laser scanning microscopy. It offers excitation with much lower laser powers than that of two-photon absorbance, but it has some potential problems because the addressing light interacts with many other data points in addition to the one being addressed.

Introduction to Confocal Microscopy - Confocal microscopy offers several advantages over conventional wide field optical microscopy, including the ability to control depth of field, elimination or reduction of background information away from the focal plane (that leads to image degradation), and the capability to collect serial optical sections from thick specimens. The basic key to the confocal approach is the use of spatial filtering techniques to eliminate out-of-focus light or glare in specimens whose thickness exceeds the immediate plane of focus. There has been a tremendous explosion in the popularity of confocal microscopy in recent years, due in part to the relative ease with which extremely high-quality images can be obtained. In fact, confocal technology is proving to be one of the most important advances ever achieved in optical microscopy.



Confocal Microscope Scanning Systems - Confocal imaging relies upon the sequential collection of light from spatially filtered individual specimen points, followed by electronic signal processing and ultimately, the visual display as corresponding image points. The point-by-point signal collection process requires a mechanism for scanning the focused illuminating beam through the specimen volume under observation. Three principal scanning variations are commonly employed to produce confocal microscope images. Fundamentally equivalent confocal operation can be achieved by employing a laterally translating specimen stage coupled to a stationary illuminating light beam (stage scanning), a scanned light beam with a stationary stage (beam scanning), or by maintaining both the stage and light source stationary while scanning the specimen with an array of light points transmitted through apertures in a spinning Nipkow disk.

Electronic Light Detectors: Photomultipliers - In modern wide field fluorescence and laser scanning confocal optical microscopy, the collection and measurement of secondary emission gathered by the objective can be accomplished by several classes of photosensitive detectors, including photomultipliers, photodiodes, and solid-state charge-coupled devices (CCDs). In confocal microscopy, fluorescence emission is directed through a pinhole aperture positioned near the image plane to exclude light from fluorescent structures located away from the objective focal plane, thus reducing the amount of light available for image formation. As a result, the exceedingly low light levels most often encountered in confocal microscopy necessitate the use of highly sensitive photon detectors that do not require spatial discrimination, but instead respond very quickly with a high level of sensitivity to a continuous flux of varying light intensity.

3. **Phase Contrast Technique.** This method usually employs a phase contrast microscope. No absorption of light is necessary, so there is no risk of damaging data while reading, but the required refractive index mismatch in the disc may limit the thickness (i.e. number of data layers) that the media can reach due to the accumulated random wave front errors that destroy the focused spot quality. As light travels through a medium other than vacuum, interaction with this medium causes its amplitude and phase to change in a way which depends on properties of the medium. Changes in amplitude give rise to familiar absorption of light which gives rise to colors when it is wavelength dependent. The human eye measures only the energy of light arriving on the retina, so changes in phase are not easily observed, yet often these changes in phase carry a large amount of information. The same holds in a typical microscope, i.e., although the phase variations introduced by the sample are preserved by the instrument (at least in the limit of the perfect imaging instrument) this information is lost in the process

which measures the light. In order to make phase variations observable, it is necessary to combine the light passing through the sample with a reference so that the resulting interference reveals the phase structure of the sample. This was first realized by Frits Zernike during his study of diffraction gratings. During these studies he appreciated both that it is necessary to interfere with a reference beam, and that that to maximize the contrast achieved with the technique, it is necessary to introduce a phase shift to this reference so that the no-phase-change condition gives rise to completely destructive interference. He later realized that the same technique can be applied to optical microscopy. The necessary phase shift is introduced by rings etched accurately onto glass plates so that they introduce the required phase shift when inserted into the optical path of the microscope. When in use, this technique allows phase of the light passing through the object under study to be inferred from the intensity of the image produced by the microscope. This is the phase-contrast technique.

2.4.3 Media Design

The active part of 3D optical storage media is usually an organic polymer either doped or grafted with the photo chemically active species. Alternatively, crystalline and sol-gel materials have been used.

1. Media form factor

Media for 3D optical data storage have been suggested in several form factors:

a. DISC. A disc media offers a progression from CD/DVD, and allows reading and writing to be carried out by the familiar spinning disc method.

b. CARD. A credit card form factor media is attractive from the point of view of portability and convenience, but would be of a lower capacity than a disc.

c. CRYSTAL or Cube. Several scientists have suggested of small solids that store massive amounts of information, and at least in principle this could be achieved with 3D optical data storage.

2. MEDIA manufacturing

The simplest method of manufacturing - the molding of a disk in one piece - is a possibility for some systems. A more complex method of media manufacturing is for the media to be constructed layer by layer.

This is required if the data is to be physically created during manufacture. However, layer-by-layer construction need not mean the sandwiching of many layers together. Another alternative is to create the medium in a form analogous to a roll of adhesive tape. Different methods such as hot stamping and photo polymerization are used in the creation of 3D data storing discs such as fluorescent multilayer disc

2.4.4 DRIVE DESIGN

A drive designed to read and write to 3D optical data storage media may have a lot in common with CD/DVD drives, particularly if the form factor and data structure of the media is similar to that of CD or DVD. However, there are a number of notable differences that must be taken into account when designing such a drive, including:

1. **LASER.** Particularly when 2-photon absorption is utilized, high-powered lasers may be required that can be bulky, difficult to cool, and pose safety concerns. Existing optical drives utilize continuous wave diode lasers operating at 780 nm, 658 nm, or 405 nm. 3D optical storage drives may require solid-state lasers or pulsed lasers, and several examples use wavelengths easily available by these technologies, such as 532 nm (green). These larger lasers can be difficult to integrate into the read/write head of the optical drive.
2. **VARIABLE Spherical Aberration Correction**. Because the system must address different depths in the medium, and at different depths the spherical aberration induced in the wave front is different, a method is required to dynamically account for these differences. Many possible methods exist that include optical elements that swap in and out of the optical path, moving elements, and adaptive optics.
3. **DETECTION.** The detection system is very different from that in a CD or DVD, and requires operation with much lower signals. When fluorescence is used for reading, special light collection optics may be used to maximize the signal.
4. **Data Tracking**. Once they are identified along the z-axis, individual layers of DVD-like data may be accessed and tracked in similar ways to DVD discs. The possibility of using parallel or page-based addressing has also been demonstrated. This allows much faster data transfer rates, but requires the additional complexity

of spatial light modulators, signal imaging, more powerful lasers, and more complex data handling.

2.4.5 Development Issues

Despite the highly attractive nature of 3D optical data storage, the development of commercial products has taken a significant length of time. This is the result of the limited financial backing that 3D optical storage ventures have received, as well as technical issues including:

1. **Destructive reading**: Since both the reading and the writing of data are carried out with laser beams, there is a potential for the reading process to cause a small amount of writing. In this case, the repeated reading of data may eventually serve to erase it (this also happens in phase change materials used in some DVDs). This issue has been addressed by many approaches, such as the use of different absorption bands for each process (reading and writing), or the use of a reading method that does not involve the absorption of energy.
2. **Stability**: Many chemical reactions that appear not to take place in fact happen very slowly. In addition, many reactions that appear to have happened can slowly reverse themselves. Since most 3D media are based on chemical reactions, there is therefore a risk that either the unwritten points will slowly become written or that the written points will slowly revert to being unwritten. This issue is particularly serious for the spiropyrans, but extensive research was conducted to find more stable chromophores for 3D memories.
3. **LASER**: As we have noted, 2-photon absorption is a weak phenomenon, and therefore high power lasers are usually required to produce it. Researchers typically use Ti-sapphire lasers or Nd: YAG lasers to achieve excitation, but these instruments are not suitable for use in consumer products.

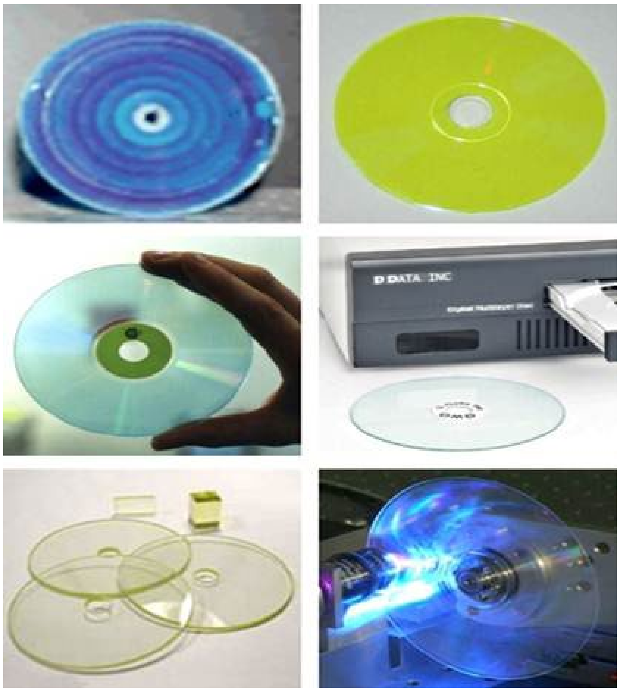
2.4.6 Commercial development

1. Flourescent Multilayer Disc

It is an optical disc format developed by Constellation 3D that uses fluorescent, rather than reflective materials to store data. Reflective disc formats (such as CD and DVD) have a practical limitation of about two layers, primarily due to interference, scatter, and inter-layer cross talk. However, the use of fluorescence allows FMDs to have up to 100 layers. These extra layers allow FMDs to have capacities up to a terabyte, while maintaining the same physical size of traditional optical discs.

Operating principles

The pits in an FMD are filled with fluorescent material. When coherent light from the laser strikes a pit the material glows, giving off incoherent light of a different wavelength. Since FMDs are clear, this light is able to travel through many layers unimpeded. The clear discs, combined with the ability to filter out laser light (based on wavelength and coherence) yield a much greater signal-to-noise ratio than reflective media. This is what allows FMDs to have many layers. The main limitation on the number of layers in a FMD is the overall thickness of the disc. A 50 GB prototype disc was demonstrated at the COMDEX industry show in November 2000. First generation FMDs were to use 650 nm red lasers, yielding roughly 140 GB per disc. Second and third generation FMDs were to use 405 nm blue lasers, giving capacities of up to a terabyte.



2. Tapestry Media

Tapestry Media is a digital optical disc about the size of a DVD with a capacity of 300GB. It will go on sale in 2009, according to its American developer, InPhase Technologies.

Traditional DVDs record data by measuring microscopic ridges on the surface of a spinning disc. Two competing successors to the DVD format — Blu-ray Disc and HD DVD — use the same technique, but exploit shorter wavelengths of light to fit more information onto the surface.

The Tapestry system uses micro holography that is light from a single laser split into two beams: the signal beam and the reference beam. The hologram is formed where these two beams intersect in the recording medium.

The process for encoding data onto the signal beam is accomplished by a device called a spatial light modulator, which translates the electronic data of 0s and 1s into an optical "checkerboard" pattern of light and dark pixels. The data is arranged in an array or "page" of around a million bits.

At the point of intersection of the reference beam and the signal beam, the hologram is recorded in the light sensitive storage medium. A chemical reaction occurs in the medium when the bright elements of the signal beam intersect the reference beam, causing the hologram.

By varying the reference beam angle, wavelength or media position many different holograms can be recorded in the same volume of material. Tapestry media is capable of storing up to 1.6TB with a data transfer rate of 120 MB/s (960 Mbit).

3. Teradisc

Mempile, a leader in next generation optical storage technology, announced today that it has proven its TeraDisc technology to be capable of storing up to one Terabyte (TB) of data. The company recently demonstrated this concept to several Japanese CE manufacturers by recording and reading over 100 virtual layers on a single DVD-size optical disc.

The demonstration attendees were amazed to see this breakthrough which showed Mempile's capability of recording at least 500GB of data on what appears to be a simple plastic transparent disc – 300GB more than the announced roadmap of competing blue-laser technologies in the year 2010.

Existing optical media store the data through the use of light-reflective semi-transparent technologies. While increasing in capacity, even the newer blue-laser technologies are nonetheless limited to a very small number of layers. The partial reflection from the multiple layers leads to signal reduction simultaneously raising background noise and coherent interferences.

Mempile's patented non-linear two-photon technology allows for 3D recording of transparent virtual layers on the entire volume of the disc. Mempile's recent demonstration proved that more than 100 layers could be recorded and read – showing storage capabilities of slightly less than 300GB over a thickness of 0.6 mm of active material. By increasing this active material to the thickness of a DVD, 1.2 mm, Mempile will be able to demonstrate the recording and reading of at least 500GB of data. Future optimization will allow the recording of 200 layers and of up to 5GB of data per layer.

Due to the increase in data retention and compliance requirements, there is also a growing need for very reliable, removable and cost-effective storage solutions such as Mempile's in the healthcare, financial, government and enterprise vertical markets. Each of these sectors now require archival storage technologies that can hold a high-capacity of information, are secure, user-friendly and are permanent yet removable and affordable. Mempile's technology is easily integrated into existing hardware manufacturing and software design processes making it a natural fit for these markets.

A Mempile disc contains light sensitive molecules (chromophores) capable of switching between two distinct states upon the application of light. Due to the nonlinear nature of the light-matter interaction, when focusing the applied light inside the material using a lens, only those molecules present near the focal point will interact and switch state. This provides for true three-dimensional accessing of small volumes within the material, allowing the writing of data bits selectively within the bulk of the material. Reading is performed in a similar way, where light that does not result in writing excites the chromophores making them emit light. The amount of light emitted is highly sensitive to there being "written" or "unwritten" molecules near the focal point, allowing this process to be used as a reading mechanism.

4. Versatile Multilayer Disc

High Definition Versatile Multilayer Disc's or Versatile Multilayer Disc (VMD or HD VMD) is a high-capacity red laser optical disc technology designed by New Medium Enterprises, Inc. VMD is intended to compete with the blue laser HD DVD and Blu-ray Disc formats and has an initial capacity of 20 GB to 40 GB per disc

Although initial details are sketchy, it appears that the format uses 5 GB per layer, similar to standard DVDs. The larger formats come from adding more layers. Whereas DVDs hold up to 2 layers per side, standard VMD's can use 4 layers, for 20 GB of storage. There are also reports of 8- and 10-layered versions which can hold 40 and 50 GB, respectively.

5. Stacked Volumetric Optical Disc

The Stacked Volumetric Optical Disk (or SVOD) is an optical disk format developed by Hitachi/Maxell, which uses an array of wafer-thin optical disks to allow data storage of around 1TB. Each "wafer" (a thin polycarbonate disk) holds around 9.4GB of information, and the wafers are stacked in layers of 100 or so, giving overall data storage increase of 100x or more. SVOD will likely be a candidate, along with HVDs, to be the next-generation optical disk standard.

2.4.7 Advantages

- A high definition movie requires about 13 GB of storage with compression so it can fit in a single disc, and there is enough space to add some extra contents such as out-takes, additional scenes, etc.
- Enables dramatic improvements in piracy protection, by taking advantage of the multiple layers of information.
- Highest optical capacity

- Lowest cost per gigabyte
- Highest data bit density of any storage device
- Lowest power requirements per gigabyte
- Long storage life
- Have highest data transfer potential

SUMMARY

- Optical storage is the storage of data on an optically readable medium. Data is recorded by making marks in a pattern that can be read back with the aid of light, usually a beam of laser light precisely focused on a spinning optical disc.
- Optical storage is any storage type in which data is written and read with a laser. Typically, data is written to optical media such as compact discs (CDs) and digital versatile discs (DVDs).
- Optical storage on the system provides an economical and efficient way to store and retrieve large amounts of information at a high performance level.
- Optical discs for data storage have different requirements. Unlike prerecorded discs that are written to only once, optical discs for storage typically need to be rewritable, which requires different recording material.
- Optical storage devices save data as patterns of dots that can be read using light. A laser beam is the usual light source.
- Read-only optical media Optical storage media that cannot be written by the user but that carry data imprinted during manufacture, usually by pressing from a master disk.
- Digital Versatile Disc - Read-Only Memory (DVD-ROM) discs can hold around 4.7GB of data (a dual-layer DVD can hold twice that). DVD-ROMs are random-access devices.
- Blu-Ray disks are a recent replacement for DVDs. A Blu-Ray disc can hold 25 - 50GB of data (a dual-layer Blu-Ray disc can hold twice that). Blu-Ray discs are random-access devices.
- HD-DVD discs are used in the same way as DVD-ROMs (see above) but, since they can hold more data, they are also used to store very high-quality, high-definition (HD) video.
- Recordable optical discs can have data written onto them (‘burnt’) by a computer user using a special disc drive (a disc ‘burner’).
- CD-Recordable (CD-R) and DVD-recordable (DVD-R) discs can have data burnt onto them, but not erased. You can keep adding data until the disc is full, but you cannot remove any data or re-use a full disc.
- The media format is the file system architecture that exists on the media to manage file, directory, and volume information.

REFERENCES

1. Costa, Dan (June 15, 2006). "Samsung Ships the First Blu-ray Player". PCMag.com. Retrieved October 17, 2007.
2. Gu, Min; Li, Xiangping; Cao, Yaoyu (May 2014). "Optical storage arrays: a perspective for future big data storage". Light: Science & Applications. 3 (5): e177.
3. Spurgeon, Brad (December 11, 2004). "Spinning out of control: Risk of CD explosion". The New York Times. Archived from the original on March 25, 2014.
4. Taylor, Jim H.; Johnson, Mark R.; Crawford, Charles G. (2006). DVD Demystified. McGraw-Hill Professional. pp. 7–8.
5. Thomann, Ralph (2005). "repair cd-player repairing cd-player cd-player help adjust cd-player repairing manual". www.ralph-toman.de. Retrieved 2020-07-31.

CHAPTER 3

OPTICAL MEMORY

INTRODUCTION

Optical memory is an electronic storage medium that uses a laser beam to store and retrieve the data. If we classify the memory system then optical memory comes under the external memory in the computer system.

Optical memory was developed by Philips and Sony and released in 1982 in the fourth generation of computers. These memories use light beams for its operations and require optical drives for its operations. These memories are used for storing audio/video, backup as well as caring for data. Read/ write speed is slower compared to hard disk and flash memories. Examples of optical memories are Compact Disk (CD), Digital Versatile Disk (DVD), and Bluray Disk (BD).

3.1 OVERVIEW OF OPTICAL MEMORY

The optical memory is a round disk with a shiny surface on which data is imprinted by means of a laser beam. Thus optical technology was first used to represent analog sound signals into digital form. The direct storage of Data as bits in memory using optical systems and properties. The memory makes use of a Laser beam that is divided by a beam splitter and controlled area of storage in memory. On the other side of the memory plane, a laser and a deflector read the memory, bit by bit, with a scanning photodetector. Erasure is accomplished by writing with the beam at a different wavelength.

3.1.1 How to use Optical Memory

The use of optical memory disks as elements in optical information processing architectures. The optical disk is an optical memory device with a storage capacity approaching 1010 bits which is naturally suited to parallel access. We discuss optical disk characteristics which are important in optical computing systems such as contrast, diffraction efficiency, and phase uniformity.

Memory is a device or system that is used to store information for immediate use in a computer or related computer hardware and digital electronic devices. The term memory is often synonymous with the term primary storage or main memory.

Computer memory operates at a high speed compared to storage that is slower but offers higher capacities. If needed, contents of the computer memory can be transferred to storage; a common way of doing this is through a memory management technique called *virtual memory*.

Modern memory is implemented as semiconductor memory, where data is stored within memory cells built from MOS transistors and other components on an integrated circuit. There are two main kinds of semiconductor memory, volatile and non-volatile. Examples of non-volatile memory are flash memory and ROM, PROM, EPROM and EEPROM memory. Examples of volatile memory are dynamic random-access memory (DRAM) used for primary storage, and static random-access memory (SRAM) used for CPU cache.

3.1.2 Laser Vision

Laser vision correction is the world's most popular elective surgery with roughly 700,000 LASIK procedures performed in the U.S. each year (AAO, 2008). Since refractive errors affect half of the U.S. population 20 years of age and older, it comes as no surprise that many people are turning to laser vision correction to obtain improved vision. Due to its popularity, medical students will inevitably be asked by patients, family, and friends about refractive eye surgery. It is important to have a basic understanding of laser vision correction, outcomes, and associated risks.

The goal of laser vision correction is to decrease dependence on glasses and contact lenses by focusing light more effectively on the retina. While there are a number of different surgeries used to achieve this result, this tutorial will focus specifically on laser vision correction, which consists of laser in situ keratomileusis (LASIK) and photorefractive keratectomy (PRK). In the U.S., LASIK comprises about 85% of the laser vision correction market with PRK making up the other 15% (ISRS). The cost of surgery varies in price from hundreds to thousands of dollars and is not covered by insurance, similar to cosmetic surgery.

Laser vision correction is regarded as highly effective with studies showing 94% of patients achieving uncorrected visual acuity of 20/40 or better at 12 months, which is the visual acuity needed to drive without corrective lenses in most states. Smaller studies have shown that up to 93% of patients with lower refractive errors can obtain 20/20 vision without correction. A meta-analysis completed in 2008 revealed that 95% of LASIK patients are satisfied with their improved vision.

3.2 TYPES OF OPTICAL MEMORY

Optical memory can be classified into many types.

Optical Memory

3.2.1 Compact Disk (CD) Technology

The first practical application of this optical technology is a compact disk i.e.CD. The compact disk is the nonerasable disk and data is imprinted on the disk using a laser beam. Initially, the CDs were designed to hold audio information from 60-75 minutes that can store about 3 GB of data. Since then the development of low cost & higher-capacity devices started.

The Compact Disc popularly known as CD is the data storage device used to store and retrieve datas encoded in digital format. It is an optical disc made of a special polycarbonate plastic known as Polymethyle Meta Acrylic. The CD reading and writing uses Laser beam generating from the Laser diode of the CD drive. Using the 780 nanometer near infrared Laser beam, the data can be stored in the Bits and Pits of the CD. Based on the Reflection and Scattering of the Laser light, The Optoelectronic tracking module in the CD drive retrieve the data.

The CD weighs 15 – 20 grams and has a thickness of 1.2 mm. Its reflecting side is coated with a thin layer of Aluminium and protected with a layer of Lacquer. Standard CD has a diameter of 120 mm and can store 700 MB uncompressed data. This can run around 80 minutes. Mini CD and Digital Video CD are also available to suit the requirements of the user.

3.2.2 The History of the CD - Technology

The development of the Compact Disc was first made possible by the invention of the laser diode, which is an essential part of the Compact

Disc and all other optical recording systems. The basic principle is that a fine laser beam is focused on a surface that contains digital information in the form of tiny pits. Since the surface of the disc is reflective, the laser beam is reflected with the pattern of the pits to a photodiode, after which the signal can be detected and converted into analogue audio information. This means there is a non-contact readout system, which cannot damage the information carrier, so that a Compact Disc in principle has an unlimited lifetime. This form of readout is highly reliable, even if the CD is worn or damaged. The CD also gives excellent reproduction quality, with negligibly low wow and flutter, a high signal-to-noise ratio and a wide dynamic range. In terms of performance the switch from analogue to digital audio processing is a more far reaching one in the history of the gramophone record than that from mechanical to electronic recording and reproduction at the beginning of the last century.

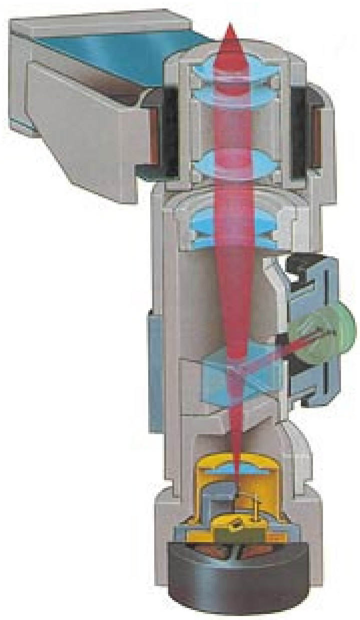
While analogue technology allows a signal-to-noise ratio of only 60 dB or less, and a low channel separation of less than 30 dB, the Compact Disc offers a much higher performance. The digital signal processing means that both the signal-to-noise ratio and channel separation are higher than 90 dB. Because a 1.2 mm thick transparent layer protects the digital information on the Compact Disc, damage and dust are not in the focal plane of the laser, which is used to 'read' the disc, so that they have little influence on reproduction. And most of the faults that arise can be corrected, thanks to the digital signal processing. This is possible because the information stored on the disc also contains error correction bits. If there are so many errors that correction is no longer possible, these can still be detected and 'masked' according to a defined procedure. So the chance that consumers will hear any of the clicks that are so well known from LPs is virtually eliminated.

The high information density means that a playing time of originally around an hour, and now up to 80 minutes, can be achieved on a disc with an outer diameter of only 12 cm. And because the disc itself is so small, the players can also be very compact. An additional feature is that a CD can also contain 'control and display' information in the form of so called C&D bits. These allow user information to be added such as the number of tracks and the playing time, as well as the names of composers and the titles of the tracks (CD Text).

The information density of a Compact Disc is relatively high, and is related to the wavelength of the laser, which is used, as well as other parameters. To allow as much information as possible to be stored, it is therefore important to use a laser with the shortest possible wavelength.

When the Compact Disc was developed, the available infra-red lasers had a wavelength of 780 nm (nanometer). The red lasers with even shorter wavelengths (650 to 635 nm), like those now used for DVD and other systems, were then not yet available.

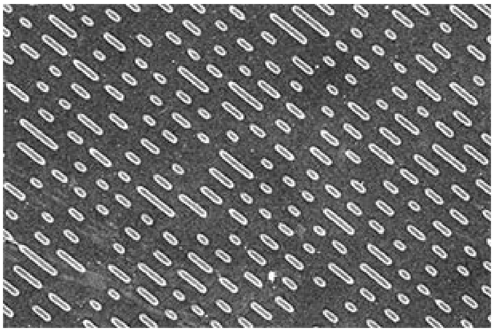
The pits in a CD are 0.6 micron wide (1 micron is 1/1000th of a mm), 0.12 micron deep and 0.9 to 3.3 micron long. A disc full of these pits, with a track pitch of 1.6 micron, has a capacity of 650 to 700 MByte. Since the information is recorded on a spiral track, and is read out at a constant speed of 1.25 m/s from the inside to the outside of the disc, the rotational speed decreases as the disc is played from 500 to 200 r.p.m.



An artist's impression of the optical pick-up, whose actual dimensions are only 45 x 12 mm.

Oversampling

With the use of oversampling technology - which allowed the 16-bit Red Book standard to be met with the 14 bit digital-to-analogue converters that were available when the system was introduced - a very high sound quality was obtained. The only difference, in the Philips players, is that this is temporarily multiplied by a factor of four. As well as allowing 16 bit-equivalent performance, this also offers additional benefits. For example it gives a higher signal-to-noise ratio and dynamic range. And instead of requiring a steep analogue output filter, it allows a relatively simple analogue filter to be used to suppress the remaining interference signals. Because even digital filters still allow some interference signals to pass through, which is why every player with digital filters also has an analogue output filter. The higher the degree of oversampling, the more effectively that a digital filter can suppress the interference signals, and the simpler the analogue output filter can be. Which in turn benefits the final reproduction quality. It's also a positive factor that the use of a digital filter has no audible effects, while a steep analogue filter causes phase changes, which affect the overall reproduction quality.



The information on the Compact Disc is recorded in digital form as a spiral track consisting of a succession of pits.

However Philips' oversampling technology, originally born out of the necessity to use the early 14 bit D/A converters, and dismissed as a 'technical joke' by other manufacturers who believed that a true 16 bit D/A converter followed by a steep analogue filter was the only way to go, was quickly embraced by most manufacturers of CD players. Because it meant there was no need to use highly complex analogue filters, while at the same time it allowed the often serious non-linearities of the D/A converters that were available at the time to be concealed.

Conversion

The conversion of the digital 'zeros' and 'ones' into an analogue signal also proved to be a tougher challenge than was at first thought. And it was also very difficult to keep the conversion process linear at lower signal levels, for example between -60 dB and -100 dB.At the introduction of the CD player, every player had a so called 'ladder' D/A converter, followed by a steep analogue filter to remove frequencies above 20 kHz. Philips was the only company to use four times oversampling, with a digital filter, right from its first player. Because four times oversampling means that four samples are taken every 1/44,100th of a second instead of just one, this in combination with first-order noise shaping, which Philips was also the first to apply, allowed 16 bit resolution to be achieved with a 14 bit D/A converter.

CD may be Recordable CD and Rewritable CD

Recordable CD

It is injection molded with a "blank" data spiral. A photosensitive dye is then applied, after which the discs are Metalized and lacquer-coated. The write laser of the CD recorder changes the color of the dye to allow the read laser of a standard CD player to see the data, just as it would with a standard stamped disc. The resulting discs can be read by most CD-ROM drives and played in most audio CD players.

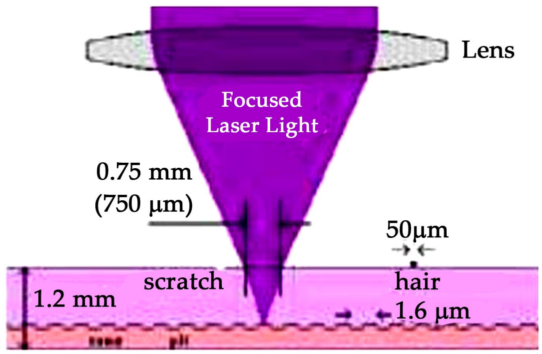
Re Writable CD

Rewritable CD is designed for 'Write and Erase' functions so that it is easy to erase the old data and record new data over it. Unlike the readable CD, the Rewritable CD has a Phase change compound coated in its reflective surface. This phase change substance is the compound of silver, antimony, tellurium and indium. This compound changes its physical state on heating depending on the temperature applied. When the temperature rises above its melting point (around 600°) it changes to liquid form and in its crystallization point (around 200°) it becomes solid. Unlike the ordinary CD, in Rewritable CD, the bumps are represented by the phase changes in the compound. When it is in the 'Crystalline form', it remains translucent so that laser light can reflect back. When the compound becomes 'Amorphous' due melting, it becomes non translucent and laser light will not reflect back. These changes during melting can 'lock' the phase change in place

The erasing process changes the crystalline and amorphous states of the compound through melting. The high temperature from the Laser beam will change the states of the compound so that the data in the form of translucent and non-translucent areas in the reflective layer will be erased. During writing, the high power laser beam melts points corresponding to the bumps of the conventional CD. They block the read laser so that it will not reflect. These non-reflective melted points remain as opaque and represent 0s. The crystalline reflective areas represent 1s.

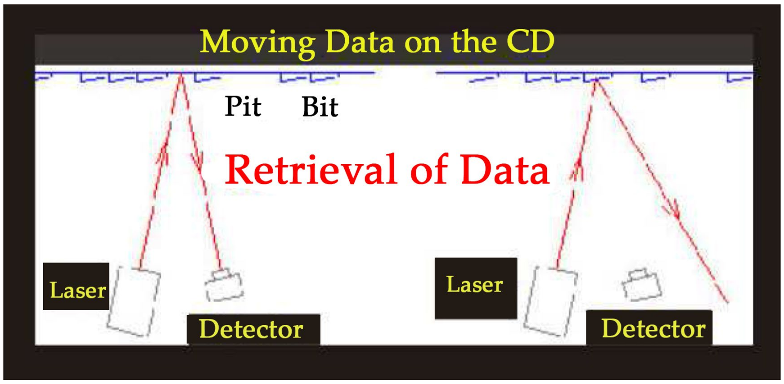
Storage Capacity

The CD can store large quantity of data as series of tiny indentations known as "pits", encoded in a spiral track molded into the top of the polycarbonate layer. The areas between pits are known as "lands". Each pit is approximately 100 nm deep by 500 nm wide, and varies from 850 nm to 3.5 μm in length. CD-ROM capacities are normally expressed with binary prefixes, subtracting the space used for error correction data. A standard 120 mm, 700 MB CD-ROM can actually hold about 737 M. Scanning velocity of the CD is 1.2–1.4 m/s (constant linear velocity) – equivalent to approximately 500 rpm at the inside of the disc, and approximately 200 rpm at the outside edge. A disc played from beginning to end slows down during playback.



CD Reading

CD players use laser technology to read the optically recorded data in the form of Bits and Pits on a CD. About 20000 or more tracks are found in a CD's recording surface. The distance between the tracks, the pitch, is 1.6 μm. A CD is read by focusing a 780 nm wavelength (near infrared) semiconductor laser through the bottom of the polycarbonate layer. The change in height between pits and lands results in a difference in intensity in the light reflected. By measuring the intensity change with a photodiode, the data can be read from the disc. The digital information is defined as the length of pits and distance between them. The pits and reflective surface represents logic 0 and logic 1. The pits and lands themselves do not directly represent the zeros and ones of binary data. Instead, Nonreturn-to-zero, inverted (NRZI) encoding is used: a change from pit to land or land to pit indicates a one, while no change indicates a series of zeros. There must be at least two and no more than ten zeros between each one, which is defined by the length of the pit.



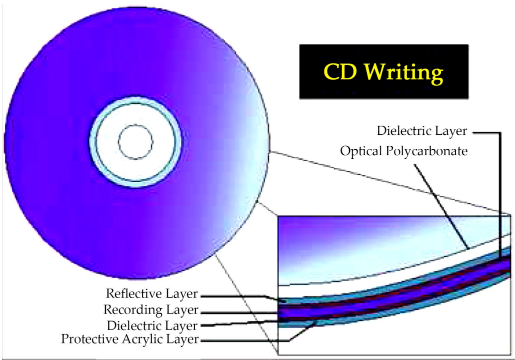
The Magic of Eye

The laser diode- lens assembly forms the optical system of the CD player. The laser diode- lens assembly is generally known as 'Eye of CD player". The lens system focuses the laser beam reflected onto the CD and reflected back light is collected by the objective lens and transmitted to the detector system. When a Laser beam is focused on to the CD, because of a difference between the depth of pits and wavelength of the laser beam, a phase difference develops between the light reflected from pits and the reflecting surface. The reflected light is then modulated by the receiver system. Before passing to the detector, the reflected laser beam is polarized and aligned to 90°. The detector is a photo sensor that produces corresponding electrical signals which are then amplified and separated into corresponding video and audio signals.

CD Burning

CD burning is an amazing technology and its working principle is really fascinating. CD burners encode music and other datas on to the blank CD which can be retrieved through CD reading. CD stores the datas in digital form represented by a series of 1s and 0s. The 1s are represented by tiny bumps and the 0s as flat areas. Such millions of bumps and flats areas are present in a standard size CD.

CD-R recordings are designed to be permanent. Over time the dye's physical characteristics may change, however, causing read errors and data loss until the reading device cannot recover with error correction methods. The design life is from 20 to 100 years, depending on the quality of the discs, the quality of the writing drive, and storage conditions. The bumps and flat areas are present on the reflective side of the CD which is arranged in continuous tracks. These tracks measures about 0.5 microns and can stretch about 5 Kms. During CD reading, a 'Read laser beam' passes over the flat area in the track. The laser beam will reflect back which will be passed onto a photo sensor assembly. The Photo sensor interprets the reflected laser light as 1. When the laser light bounce back from the bumps, the photo sensor will not get it and the CD player recognize it as 0.



CD writer is used to burn the CD to record the data in the digital format. The CD burner darkens microscopic areas in the reflective side of the 'Blank CD'. During burning, both reflective and non-reflective areas are created in the CD that can be interpreted as 1 and 0 by the CD player. The CD burner has a laser assembly similar to the CD reader. But the laser is 'Write Laser' with high power around 40 mW. The write laser interacts with the CD and alters its surface. The movement of the Write laser assembly is exactly similar to the read laser assembly. It moves outward from the center of the CD while the CD is spinning. The bottom layer of the 'Blank CD' has grooves pre pressed into it to guide the write laser in the correct path. By calibrating the spin speed and speed of the laser assembly, the burner will guide the laser assembly exactly through the track at a constant speed. During the movement, the laser diode will turn on and off in synch with the pattern of 1s and 0s in the track. The high power laser then darkens the CD material to encode 0 and leaves the material translucent to represent 1.

The CD writing speed can be varied depending on the choice. The CD burner can write at multiple speeds. 1x is the writing speed exactly similar to the reading speed. That is ,it takes 60 minutes to write 60 minute duration data. At 2x rate, the burner will take 30 minute to write 60 minutes data. The general writing speeds are 8x, 16x, 24x, 32x etc. It is better to avoid high speed writing since at high speed, some of the data will not be stored especially when writing Music files. The normal writing speed is 16x or 24x.

CD Writing Software

To control all the functions of the CD writer, software like Nero is used. The datas accessible for all types of CD players are encoded in the understandable form by the software. The programme also reduces the data errors during writing. This is achieved by incorporating a number of extra digital information with the recorded data and arranges them carefully. The format gives 'Time Codes' during writing so that the CD player can recognize which part is reading at a particular time. The format also gives 'Table of contents' at the beginning of the track (center of CD) so that the CD player can recognize, which file is reading. 'Extra data bits' are also added so that the player can fix the mistakes if the laser misread the data bit. The recorded data is not arranged sequentially but in an 'inter laced pattern'. This prevents the loss of complete data from the CD, if a portion of the CD is damaged

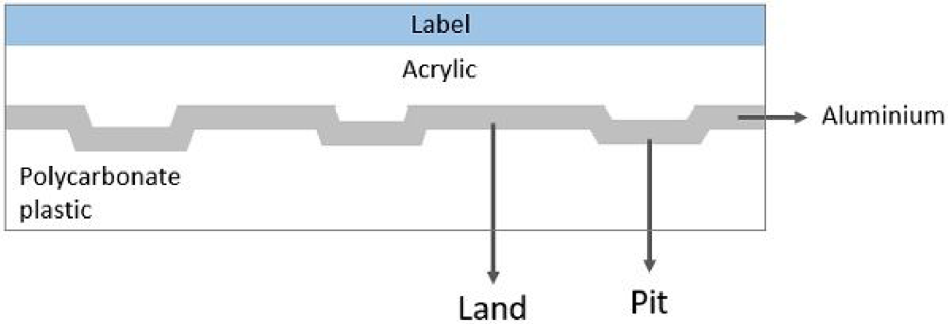
CD damage

CDs are susceptible to damage from both daily use and environmental exposure. Pits are much closer to the label side of a disc, so that defects and dirt on the clear side can be out of focus during playback. Consequently, CDs suffer more scratch damage on the label side whereas scratches on the clear side can be repaired by refilling them with similar refractive plastic, or by careful polishing.

Constructional Structure of CD

The bottom layer of the compact disk that has exposure to the laser beam is formed by the transparent polycarbonate plastic. This appears as the clear glass base and its surface is programmed to store data. The laser beam is used to imprint data on the polycarbonate plastic by indenting it with *pits.*

The indented parts are termed *pits* whereas the parts that remain unindented are termed as *lands*. The layer of this polycarbonate plastic is covered with a thin layer of aluminium which makes it shiny. To prevent this shiny surface from dust and scratch, it is covered by a thin layer of acrylic.



Cross Section of Compact Disk

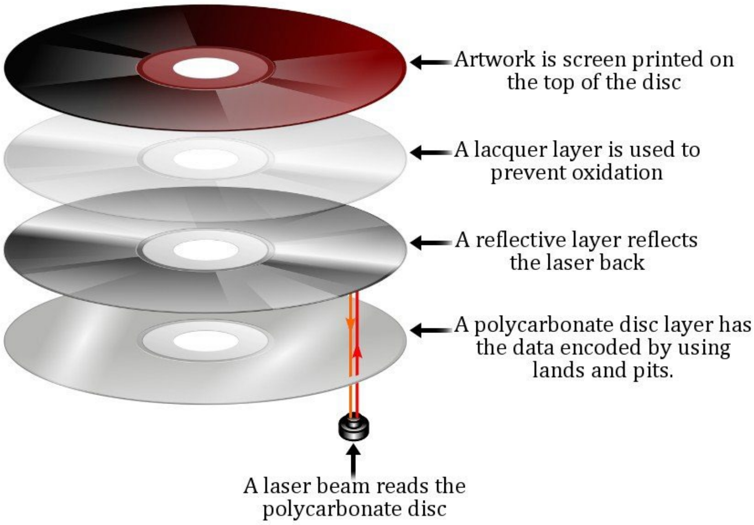
Finally, a label is stamped on to the acrylic surface. The total thickness of the CD is 1.2 mm to which the polycarbonate plastic contributes the most.

3.2.3 How Does A Compact Disc (CD) Work?

A few years ago, when USB sticks and cloud computing were not as popular as they are today, data was primarily stored and retrieved with the help of CDs. Today, shiny circular discs are known for storing data ranging from a few hundred megabytes to a few gigabytes.

How does a CD work?

A CD is usually around 12 centimeters (4.5 inches) in diameter and consists of a couple of thin circular layers attached one on top of another.



Various layers of a CD

Most of a CD is composed of a plastic called polycarbonate. The bottom layer is a polycarbonate layer where data is encoded by using tiny bumps on the surface. Above this layer is a reflective layer typically made of aluminum (gold is also used, although quite rarely).

Above the reflective layer is a protective layer of lacquer and plastic, which shields the layers below. The artwork or label is printed on the lacquer layer (i.e., on top of the CD) via offset printing or screen printing.

CDs store information digitally, i.e., with the help of millions of 1s and 0s. Data on a CD is encoded with the help of a laser beam that etches tiny indentations (or bumps, if you will) on its surface. A bump, in CD terminology, is known as a **pit** and represents the number 0. Similarly, the lack of a bump (known as **land**) represents the number 1. Hence, a laser beam can encode the required data into a compact disc using pits and lands (0 and 1, respectively). Now that you know how a CD is encoded with data let's look at how a CD player actually reads this stored data.

How does a CD player work?

There are two main components in a CD player that help read a CD: a tiny laser beam known as a semiconductor diode laser and an electronic light detector, basically a tiny photoelectric cell. When you turn on the CD player, an electric motor in the player rotates the CD at a very high speed while reading the outer edge at 200 RPM, and when reading the inner edge, it rotates at 500 rpm.

The laser beam source inside the player switches on and scans along a track from the center of the disc to the outer rim. It focuses a 780 nm wavelength (near-infrared) beam through the underside of the compact disc. When the beam falls on land (1), it reflects straight back, but it scatters when the beam falls on a pit (0).

When the photocell detects the reflected light, it recognizes that the laser must have hit land and, in turn, sends a signal to a circuit that generates the number 1. Likewise, when it does not detect light, it correctly determines a pit at this point so that the circuit generates the number 0. Thus, the photocell uses the intensity changes of the reflected beam to determine whether there is a 1 or a 0 on the disk.

All kinds of digital information can be stored with the help of 0s and 1s.

Why is it difficult for a CD player to read the contents of a scratched CD?

The data from a CD / DVD / Blu-ray disc is not on the glossy surface but the polycarbonate layer at the bottom of the disc. As already mentioned, a CD player has a laser beam that reflects/scatters from the underside depending on whether it falls on a land/pit. Indentations on the surface of a disc are very, very small, so scratches and cracks mess up the way light bounces off the surface of the CD.

When the laser falls on a scratched spot, it scatters, even if there is no bump at this point. As a result, the photocell transmits incorrect information to the circuit, making it difficult for the CD player to read the data correctly.

3.2.4 Mechanism of Compact Disk

How the digital data is imprinted on the compact disk?

The laser beam with high intensity is focused on the disk which imprints or indents the series of microscopic *pits* on the surface of polycarbonate plastic. These pits are arranged on a long circular track on the surface of the disk spiraling from the center of the disk to the outer edge thus forming a master disk.

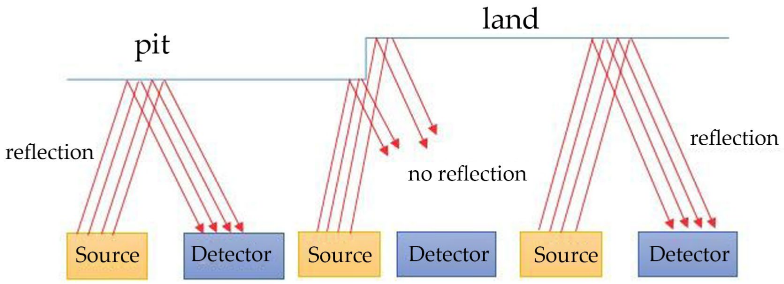
This master disk is then used to produce a mold which can be used to produce a large volume of CDs that hold the same information as that of master CDs.

How the data is retrieved from the compact disk?

The low power laser beam is focused on the polycarbonate plastic. The laser beam travels through the polycarbonate plastic layer and gets reflected back from the aluminium layer.

This reflected beam is now received by the photodetector which is also positioned below the polycarbonate plastic.

When a laser beam falls on the pit which is a slightly rough surface the intensity of the reflected beam is low. When the laser beam falls on the land surface the reflected beam return with a higher intensity.



Transition from Pit to Land

The photodetector receives the reflected beam and senses the change in the intensity of the reflected beam thereby convert it into a digital signal. As the beginning or the ending of pits when there occur a change in the elevation of the reflected beam is recorded as 1. The smooth land surface where there is no change in the elevation of the reflected beam is recorded as 0 by the photodetector.

The figure below shows you several transitions of the laser beam between lands and pits.

Detected Binary Pattern by Photodetector

CD-ROM

Compact Disk Read-Only Memory (CD-ROM) is a read-only memory that is used to store computer data. Earlier the CDs were implemented to store audios and videos but as the CDs store data in digital form they can be used to store computer data.

The audio and video appliance can tolerate some amount of error as it does not reflect in the produced audio and video in a noticeable way. But when it comes to computer data the computer appliances do not tolerate any error. Now it is not possible to prevent physical imperfection while indenting pits on the CDs. So some extra bits must be added to detect and correct the error.

The CDs and CD-ROM has a single spiral track that starts from the center of the track and spiral out toward the outer edge. The data is stored in CD-ROM in blocks i.e. *sector.* The number of sectors varies from track to track. There are fewer sectors in the inner tracks of the CD and more sectors in the outer track of the CD.

The sector at the outer and inner edge of the disk is of the same length. These sectors are scanned with the low power laser beam at the same rate while the disk is rotating. Though the rotating speed of the disk can vary. To access the sectors near the center, the disk rotates comparatively faster as compared to access the sectors present at the outer edge of the disk.

CD-Recordable

CD-Recordable i.e. (CD-R) was the first kind of compact disk that could be easily recorded by any computer user. This disk has a similar shiny spiral track as we can see in CD and CD-ROM. This shiny track is cover with *organic dye* at the time of manufacturing.

To record the data on the CD-R the disk is inserted into the CD-R drive and a laser beam is focused on the drive which burns *pits* onto the *dye.* The burned spot become opaque and the unburnt area still appears shiny. When the laser beam with low power is focused on the disk to retrieve the information. The opaque spots reflect light with less intensity and the shiny parts reflect light with high intensity.

Remember the CD-R can be burnt or recorded once in its lifetime. Though the unused portion of the CD-R can be used to record some more information later.

3.2.5 CD-Rewritable (CD-RW)

This CD can be recorded multiple times which means the user can write and erase data from the CD-RW multiple times. This is because instead of using the organic dye an alloy is used which includes silver, indium, antimony, and tellurium. The melting point of this alloy is 500° C.

The alloy shows interesting behavior when it is heated and cooled down. When the alloy is heated above the melting point and cooled down it turns into an amorphous state which is capable to absorb light.

In case the alloy is heated at 200° C and maintained at that temperature for a certain period a process *annealing* takes place which turns the alloy into the *crystalline* state. At this state the alloy allows the light to pass through it.

So here the pits can be created by heating the selected spots above the melting point and the remaining parts between the pits are lands. The stored data can further be deleted using the annealing process.

3.2.6 Digital Versatile Disk (DVD)

The DVD (common abbreviation for Digital Video Disc or Digital Versatile Disc) is a digital optical disc data storage format invented and developed in 1995 and released in late 1996. The medium can store any kind of digital data and was widely used for software and other computer files as well as video programs watched using DVD players. DVDs offer higher storage capacity than compact discs while having the same dimensions.

Prerecorded DVDs are mass-produced using molding machines that physically stamp data onto the DVD. Such discs are a form of DVD-ROM because data can only be read and not written or erased. Blank recordable DVD discs (DVD-R and DVD+R) can be recorded once using a DVD recorder and then function as a DVD-ROM. Rewritable DVDs (DVD-RW, DVD+RW, and DVD-RAM) can be recorded and erased many times.

DVDs are used in DVD-Video consumer digital video format and in DVD-Audio consumer digital audio format as well as for authoring DVD discs written in a special AVCHD format to hold high definition material (often in conjunction with AVCHD format camcorders). DVDs containing other types of information may be referred to as DVD data discs.

The DVD technology was first introduced in the year 1996 and has the same appearance as that of the CD. The difference is in their storage size, the DVD has much larger storage than that of CD and this is done by implementing several changes in the design of the DVD. The laser beam used imprint data in DVD has a shorter wavelength as compared to the wavelength of laser beam we use for CDs. The shorter wavelength of the laser beam helps the light to focus on a smaller spot.

Pits are much smaller as compared to pits of CD and even tracks are placed much closer as compared to the tracks in CD. With all these changes in design, the DVD has a storage size of 4.7 GB. To increase the storage capacity even more the two-layered and two-sided disk was introduced.

Two Layered DVD

The two-layered disk has the first base as present in CD but instead of using aluminium, the lands and pits of the first base are covers using a *translucent* material that solves the purpose of the reflector.

Now, this translucent layer is also programmed to store the data by indenting pits onto it. Reflective material is placed on this second layer of pits and lands.

When the laser beam is focused on the first layer to retrieve the binary pattern the translucent material will reflect with sufficient light to be captured by the detector. The second layer will reflect a small amount of light which will be cancelled by the detector as *noise.*

Similarly, when the laser will be focus to read the second layer the first layer will reflect with a small amount of light that would be cancelled by the detector.

Two-Sided DVD

The tracks are implemented on both sides of DVDs. This structure can be understood as the two single-sided disks that are put together to form a sandwich but the topmost disk is turned upside down.

High-Definition Optical Disks

The high definition optical disk is the advanced version of DVDs that are used to store high definition videos. To raise the capacity the laser beam with an even shorter wavelength in the blue-violet range is used. Pits on high definition optical disk are even shorter when compared to DVD.

So these all are types of optical memory that can be used to store digital data. The optical storage is durable, easy to transport and are capable of storing a vast amount of data.

3.2.7 Compact Disc Digital Audio

Compact Disc Digital Audio (CDDA or CD-DA), also known as Digital Audio Compact Disc or simply as Audio CD, is the standard format for audio compact discs. The standard is defined in the *Red Book*, one of a series of Rainbow Books (named for their binding colors) that contain the technical specifications for all CD formats.

The first commercially available audio CD player, the Sony CDP-101, was released October 1982 in Japan. The format gained worldwide acceptance in 1983–84, selling more than a million CD players in those two years, to play 22.5 million discs.

Beginning in the 2000s, CDs were increasingly being replaced by other forms of digital storage and distribution, with the result that by 2010 the number of audio CDs being sold in the U.S. had dropped about 50% from their peak; however, they remained one of the primary distribution methods for the music industry. In the 2010s, revenues from digital music services, such as iTunes, Spotify, and YouTube, matched those from physical format sales for the first time. According to the RIAA's midyear report in 2020, phonograph record revenues surpassed those of CDs for the first time since the 1980s

3.3 OPTICAL RAM AND INTEGRATED OPTICAL MEMORIES: A SURVEY

The area of integrated optical memories and optical random access memories (RAMs) together with the rapid adoption of optical interconnects in the Datacom and Computercom industries introduce a new perspective for information storage directly in the optical domain, enabling fast access times, increased bandwidth and transparent cooperation with optical interconnect lines.

Over the past decades, "storing light" has appeared as a rather controversial statement, given that a photon's inherent nature hinders its spatial confinement. The first research efforts in demonstrating optical memory functionality started as a fascinating experimental exercise, with the first optical memory being reported by means of a folded optical delay line back in 1965. Two decades later, the first optical set-reset flip-flop (SR-FF) mechanism was launched in 1985, achieving response times of <1 ns, while in the next few years, research efforts mainly focused on temporarily confining light to a continuous loop inside a medium. As fiber optics gradually turned into a mainstream telecom transmission platform, the research interest in optical memories experienced a significant boost in view of the possible high-speed optical signal processing applications, with a variety of schemes such as optical delay lines, fiber-loop-based and slow-light optical buffers and, more recently, all-optical flip-flop (AOFF) devices being introduced for packet-level contention resolution purposes.

Following the initial attempts to store light for packet-level processing, optical memories have made significant progress during the last decade and managed to penetrate the area of bit-level storage, significantly expanding along the performance metrics, functionality and application perspectives. This progress has been greatly facilitated by the rapid advances in photonic integration and the massive penetration of optics at interconnect segments closer to the CPU level. At the same time, the well-known memory-related bottlenecks in the fields of computing and routing have served as the main motivating use-cases for transferring the speed and energy advantages of light technology to the memory domain, with the CPU-memory bandwidth bottleneck and the more recent decline of Koomey's law comprising just two indicative examples driving research toward optical random access memories (RAMs) and optical memories for non-Von-Neumann computing paradigms, respectively.

Figure 1 presents an overview of the most important categories into which current optical memories can be classified. Based on the size of the data information that is stored, i.e., a data bit or a complete data packet, optical memories can be categorized in (a) bit-level and (b) packet-level configurations, with packet-level buffering performed by more conventional and older delay line and recirculating loop technologies. Similar to electronic technology, optical bit-level memories can in turn be classified as either volatile or non-volatile structures, depending on whether the stored data are lost or maintained, respectively, when the power supply is switched off. Optical volatile memories can typically offer faster access times and higher speed operation compared to their non-volatile counterparts and form the core memory mechanism in the optical versions of the well-known and highly useful RAM cell architectures, again discriminated into two main categories: (a) the optical dynamic (DRAM) and (b) the optical static (SRAM) RAM, with their main difference lying in their requirement for refreshing (DRAM) or not (SRAM) the stored bit value. Optical SRAM layouts have thus far been implemented mainly by means of bistable optical devices, whereas the optical DRAM cells that have been reported rely on either low-speed optical physical mechanisms such as ion excitation or recirculating loop arrangements. Optical non-volatile memories are a more recent addition to light-enabled memory technology, mainly taking advantage of the rapid progress experienced in the field of phase-change material (PCM) structures, which have been shown to allow for permanent light storage in a continuously growing field of diverse applications.

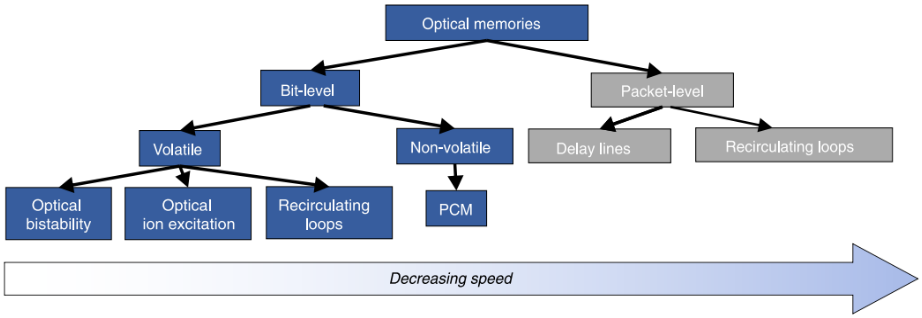


Figure 1. Classification of optical memory technologies

The substantial progress witnessed in the field of integrated optical memory technologies, mainly focusing on bit-level volatile and non-volatile optical structures and on roadmaps for transforming these elementary optical memory modules into practical optical RAM cell layouts. The paper is organized as follows: First, the basic approaches and principles applied to achieve light-based storage in general are presented, and the main technical system requirements in terms of memory are discussed. Then, the state-of-the-art optical memory technologies, and their benefits in terms of energy, bandwidth and footprint are summarized. Following this, advanced memory functionality i.e., true optical RAM operation, is explained, and recent advancements are reported. Finally, an analysis is presented for the next steps that optical memory technologies must undertake to release a viable and practical alternative memory roadmap.

3.3.1 How to Store Information with Light

Although light has inherent disadvantages when considering buffering functionalities, as the neutral charge of photons makes it impossible to mimic the respective capacitor-based electronic memory layouts, the research community has devised several methods to enable light-based storage. The most common approaches to achieving this rely on either the bistability of engineered optical resonances (artificial cavities), such as in refs, or the inherent bistable characteristics of devices stemming from their material properties. Two main conditions should be applied to achieve optical bistability and consequently memory operation: the system should (a) provide at least two discrete, stable states that represent the logical one and logical zero and (b) allow switching between the two states under certain conditions. Figure 2 summarizes the four most popular categories of bistable memory devices, which rely on (a) the master-slave configuration, (b) the feedback loop scheme, (c) the injection-locking technique and (d) phase-change materials. In the master-slave configuration (Fig. 2a), two active components that are usually either switches or lasers are placed in a coupled arrangement, forming an artificial cavity. In this case, the discrete memory states (i.e., logical value “1” or logical value “0”) are represented by two different states of a certain light beam characteristic, such as the polarization or wavelength of the light beams emitted by the respective active components.

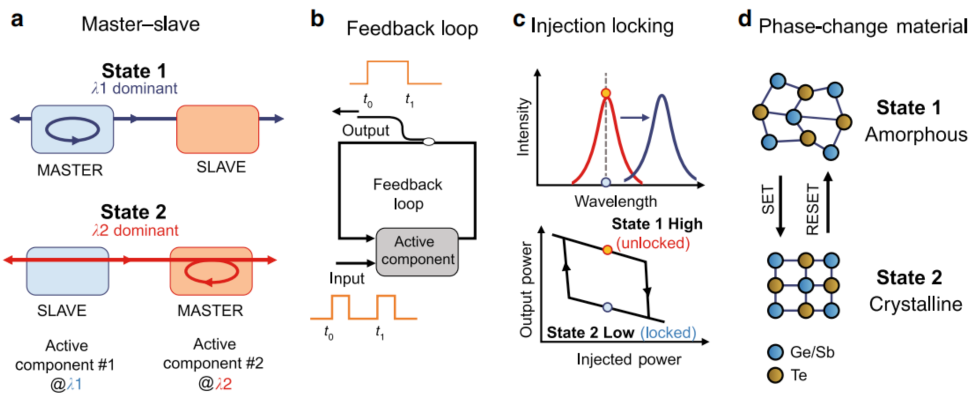


Figure 2. Optical memory bistability based on different approaches**. a** master-slave scheme, **b** feedback loop scheme, **c** injection-locking technique and **d** phase-change material (PCM) properties in the case of GST compounds

Figure 2a depicts the case where two different wavelengths are used to denote the different binary states. Each time, only one of the two available wavelengths can be dominant in the cavity, whereas the other remains suppressed. Assuming that wavelength λ_1 corresponds to the logical value of “1” and wavelength λ_2 represents the logical value “0” emitted by active component #1, State 1 refers to the cavity situation where light at λ_1 dominates the cavity and suppresses wavelength λ_2 emitted by active component #2. As long as State 1 is dominant, active component #1 serves as the “master”, whereas active component #2 is the “slave”, with the memory output signal obtained at wavelength λ_1. Conversely, in State 2, wavelength λ_2 suppresses wavelength λ_1, and the memory output emits a signal at wavelength λ_2. Changing between the two states is accomplished by injecting external light at the appropriate amount of power and wavelength into the "master" component, suppressing its operation and allowing sufficient time for the "slave" device to recover to its equilibrium state. In this case, the wavelength emitted by the "slave" device can then reach the "master" device, acting as a holding signal that retains the suppression of the former "master" wavelength even if the external light injection stops. This type of scheme is usually employed for set-reset flip-flops (SR-FFs), which have also been employed in optical SRAM cells. To date, theoretical studies on coupled schemes have revealed that the switching time between two states is inversely proportional to the length of the cavity formed between the two active components, suggesting that an integrated solution has to be adopted to enable switching times in the picosecond regime.

Optical memories based on feedback loops, shown in Fig. 2b, require a single active component along with an external cavity usually implemented by loop configurations that feed the output signal back to the active element either through a fiber or by using an integrated bent waveguide. The cavity acts as the memory element, enabling bit storage, and a tap of the cavity allows for monitoring the logical state of the feedback loop, i.e., the memory content. The active element employed so far is a 1×2 optical switch that either feeds the loop with the switched signal supporting its recirculation or blocks the recirculation by switching the signal out of the loop. This type of optical memory has been demonstrated in SR-FF schemes using independent and discrete set and reset externally injected signals but has also been employed to build toggle flip-flops (T-FFs) by applying a single external pulsed signal, as depicted in Fig. 2b. The demonstrated T-FFs follow an electronics-borrowed approach where two options are available: (a) either maintain the current state's value for another cycle in the case of a logical zero or (b) toggle the value (negate it) at the next clock edge in the case of a logical one at each input. In that case, the loop retains its state when the incoming signal is blocked; otherwise, the memory content is changed, yielding a T-FF functionality that is highly useful for shift registers and counters.

On the other hand, the injection-locking technique widely used in lasers can provide optical memory bistability by forcing specific light characteristics of the lasing device to lock to the respective characteristics of an externally injected optical beam. The light characteristics that can be applied through the locking mechanism are usually the (a) wavelength, (b) polarization state, and (c) propagation direction. The memory bistability is observed as the interchange of the laser emission states between a free-running mode (unlocked/high state) and an injection-controlled mode (injection-locked/low state). The principle of operation in the case of wavelength bistability is graphically represented in Fig. 2c. Initially, the laser emits at its free-running state signal, shown in red in Fig. 2c. In the presence of a wavelength-detuned input injection signal, called the control signal, the laser starts emitting at the injection wavelength (blue in Fig. 2c) and not at its free-running mode wavelength when the control power increases above a specific threshold. As the optical power of the control signal decreases to a certain value at this state, the device enters a hysteresis loop retaining this emission state even when the control signal optical power is decreased to a certain cut-off level. Figure 2c illustrates an indicative hysteresis loop formed by a laser device assuming a given wavelength detuning. As is evident, when the optical power of the injected signal falls below this cut-off level, the laser emission returns to its free-running "unlocked" state. Consequently, the laser emission output has two states, i.e., locked (low) and unlocked (high), which depend on the ascending or descending direction of the injection signal power, and the memory operation can be achieved when operating within the bistable range of the laser device. Similarly, memory bistability can be achieved by means of polarization by interchanging the polarization state (orthogonal or vertical polarization) of the injected optical signal, while in the case of the propagation direction, the light in the device can be forced to circulate either to the clockwise (CW) or to the anticlockwise (ACW) propagation mode by setting the injected external signal in the appropriate direction.

Another approach to enabling optical memory bistability relies on the exploitation of the physical properties of the optical phase-change materials (O-PCMs). O-PCMs have emerged as a unique class of materials that can exhibit large changes in their optical properties (index change $\Delta n > 1$, $\Delta\kappa$ ~ order of magnitude) in response to an external stimulus (i.e., temperature, applied voltage or ultra-fast optical excitation). Most established O-PCMs for optical memories are chalcogen-based alloys such as $Ge_2Sb_2Te_5$ (commonly known as GST), in which the material undergoes transitions between its amorphous and crystalline states. An example of the principle of operation of a PCM-based optical memory is shown in Fig. 2d. In this recently introduced all-optical PCM memory, a small patch of GST loaded on top of a silicon-nitride waveguide is used, and memory bistability is triggered by injecting optical pulses that can lead the thin film to adopt either an ordered crystalline or disordered amorphous state. In Fig. 2d, different colors represent different atoms, such as Ge, Sb, and Te, in the GeSbTe compound. The phase of the GST element affects the optical properties of the underlying waveguide such that the specific phase and subsequently the memory content can be concluded by monitoring the intensity of the propagating light at the output. In the crystalline state, the GST is more absorptive, inducing strong attenuation to the propagating light, which results in low intensity at the output that corresponds to the logical “0”. On the other hand, in the amorphous state, the absorption is reduced, allowing for high-intensity pulses at the output and yielding a logical value of “1” at the memory output. Switching between the two phase states occurs when high-intensity optical pulses are injected and, based on their total energy, can initiate either amorphization (write) or crystallization (erase). It is important to note that this type of memory element can also be configured to support multiple intermediate absorption levels between its two extreme states, allowing for multi-level operation and multi-bit storage properties.

3.3.2 State-of-the-art Optical Memory Technologies

The current state-of-the-art optical memory technologies. Figure 3 summarizes the most popular optical volatile memory technologies that have been successfully pursued toward delivering light-based storage, relying mainly on (a) VCSELs, (b) semiconductor optical amplifiers (SOAs), (c) InP coupled ring lasers, (d) an InP microdisk laser, (e) InP buried heterostructure (BH)-PhC nanocavity switches, and (f) hybrid InP-on-SOI PhC lasers. Figure 4 presents the non-volatile PCM optical memory technology platform and its main principle of operation.

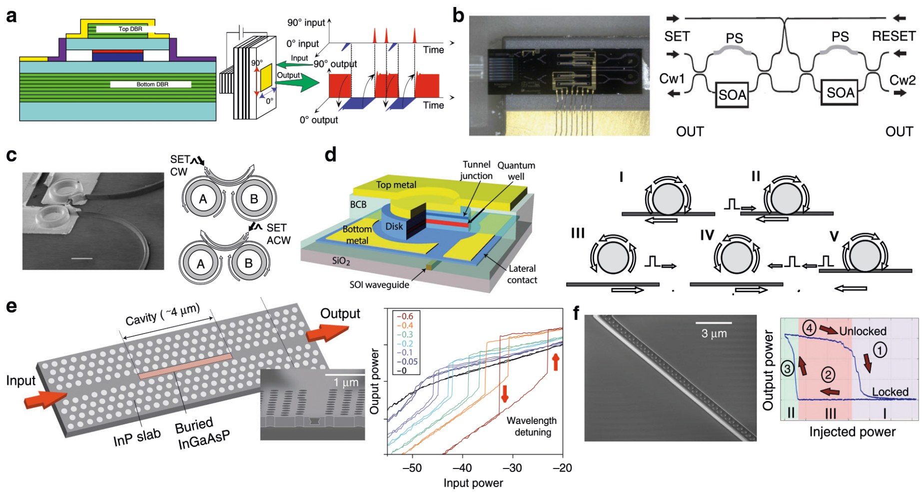


Figure 3. Optical memory devices and their respective principle of operation. a VCSEL-based optical memory, **b** an SOA-MZI coupled optical memory, **c** an InP micro-ring laser memory, **d** an InP microdisk memory, **e** an InGaAsP photonic crystal nanocavity memory and **f** an InP-on-SOI hybrid photonic crystal nanocavity laser memory.

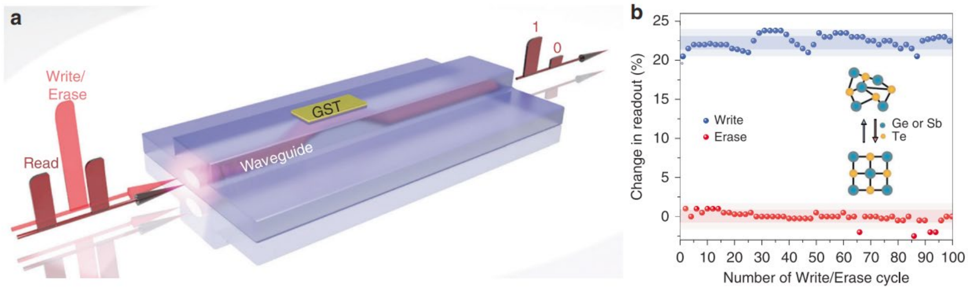


Figure 4. Optical phase-change memory device and its respective principle of operation. a Information is stored in the phase state of the GST section on top of the nanophotonic waveguide. b Demonstration of binary memory operation between the crystalline (lower, level 0) and amorphous (upper, level 1) states of a 5 μm GST device for multiple repetitions of the same switching cycle.

3.3.3 Vertical Cavity Surface Emitting Lasers (VCSELs)

The first optical memory demonstration based on a bistable vertical cavity surface emitting laser (VCSEL) was reported in 1991 58. Since then, significant research efforts have been invested toward achieving VCSEL-based memories. The VCSEL structure for polarization bistability at the 1.55 μm wavelength region and its respective principle of operation are shown in Fig. 3a. It exploits the polarization bistability, which is controlled through the injection of an external signal with an appropriate power and polarization direction (orthogonal or vertical), such that the polarization state of the VCSEL output signal follows the polarization state of the injected optical pulse. In this way, the memory content (i.e., logical state "0" or "1") of the VCSEL optical memory is identified based on the polarization state of the output signal. Polarization bistable 980 nm VCSEL-based memories have been demonstrated with 20-Gb/s RZ and 40-Gb/s NRZ optical pulses 40 at slower repetition periods, enabling multi-bit memory implementations. The main advantages of polarization-bistable VCSELs include (a) their potential for high-speed memory operation, being able to handle up to 40 Gb/s optical pulses, (b) their attractive properties for logic gate functionalities, (c) their low-energy consumption requirements (~105 fJ for 40 GHz operation) compared to other types of bistable laser diodes, stemming from their lower bias current requirements, and (d) their established and mature laser technology platform, which can form the basis of a reliable optical memory solution. Among their main limitations are certainly the relatively increased footprint requirements, as only the active square mesa corresponds to ~36 μm^2, and the need for a carefully controlled polarization state and alignment, especially when targeting multi-bit integrated modules.

3.3.4 Semiconductor Optical Amplifier (SOA)-Based Schemes

In the early 2000s, several AOFF demonstrations were introduced relying on semiconductor optical amplifier (SOA)-based lasers or switches performing in master-slave configurations and exploiting well-known non-linear physical phenomena such as gain saturation and polarization-dependent gain saturation in SOA-based switches. These demonstrations were implemented with discrete fiber-pigtailed components, with the first integrated AOFF appearing in 2006. This AOFF utilized hybrid silica-on-silicon integration technology and a coupled SOA-MZI-based architecture 68 and mainly targeted all-optical packet switching applications to facilitate routing and forwarding directly in the optical domain. Bit-level optical memory implementations were demonstrated shortly after utilizing cross-phase modulation (XPM) phenomena in SOA-MZIs, cross-gain modulation (XGM) in coupled SOAs or SOA-based coupled ring lasers. The transfer of this AOFF scheme into its InP-based monolithically integrated version, which was then also employed in true optical RAM cell setups, was only recently demonstrated, reporting 10 Gb/s operation and a drastic footprint reduction of 97.8%% compared to its hybrid-integrated predecessor. A photo of the monolithic integrated device and its principle of operation are shown in Fig. 3b, illustrating that it follows a master-slave configuration, with the two coupled SOA-MZIs being powered by two external continuous-wave (CW) input signals CW1 and CW2 and the logical value of the memory cell being determined by the wavelength of the dominant CW signal. Other AOFF schemes based on SOA-based DFBs, SOAs in combination with DFB laser diodes, loop mirror setups, and feedback loops have also been presented. Among the main benefits of SOA-based technologies in optical memory implementations are (a) their enhanced maturity level and flexibility characteristics that, in many cases, allowed for the proof-of-concept demonstration of novel memory concepts prior to proceeding to their more compact and integrated versions and (b) their high-speed potential, having already resulted in 10 Gb/s memory line rates and being theoretically predicted to allow up to 40 Gb/s operating speeds even in optical SRAM cell arrangements. However, their energy and footprint drawbacks probably critically affect its practical perspectives: SOA-based AOFFs require very large amounts of energy for both SOA biasing (~120 pJ and ~180 pJ) and for optically switching between set and reset states (~3 pJ, ~0.5 pJ), with the current footprint requirements hardly going below a few mm^2.

3.3.5 (Micro) Ring Lasers

A fast, low-power AOFF-integrated memory based on coupled micro-ring lasers exploiting the injection-locking technique was reported 30 in 2004, with its principle of operation shown in Fig. 3c. By connecting two ring lasers together via a waveguide, as depicted in Fig. 3c, two inherent lasing modes can be exploited to create a system where the master micro-laser injection locks the slave laser under certain conditions and defines the direction of the propagating light, dictating in this way two possible stable states: (a) laser light traveling in the clockwise (CW) direction and (b) laser light in the anticlockwise (ACW) direction. To switch states, light close to the lasing characteristics in terms of the wavelength and polarization needs to be injected into the waveguide connecting the lasers to set both lasers to lase simultaneously in either the CW or ACW direction. Alternative AOFFs and optical memory demonstrations relying on semiconductor ring lasers have also been suggested following the rationale that the cavity should support two counter-propagating directional modes. The first demonstration was a novel single semiconductor micro-ring laser employing a retro-reflector cavity to enable 2-bit optical storage while achieving fast ON/OFF switching times. Another semiconductor ring laser was also proposed by CNIT in 2013, who reported high-speed operation at 10 Gb/s and an improvement in the switch-ON times up to 10 ps. The proposed micro-ring lasers can provide electrically pumped optical memory implementations, also requiring, however, an additional DC current bias to tune the resonant frequencies of two lasers close to each other. Although integrated ring laser schemes can offer some attractive advantages such as (a) multi-Gb/s operational speeds (10 Gb/s 33) and fast switching times of 20 ps and (b) high-output-signal extinction ratio values that can reach almost 40 dB, their main drawbacks remain (a) the total energy consumption accounting for several pJ (~1.2 pJ and 54 pJ) and being mainly dominated by the bias current (30 mA, ~200 mA) and (b) their large footprint, occupying more than 1000 μm^2 and reaching, in some cases, even several mm^2.

3.3.6 Microdisk Laser

An ultrasmall, low-power, electrically pumped AOFF memory on a silicon chip was introduced by IMEC in 2010. The AOFF relied on a single microdisk laser with a diameter of 7.5 μm coupled to a silicon-on-insulator (SOI) wire waveguide. Figure 3d shows a schematic of the microdisk laser and its principle of operation, which again exploits the propagating light direction to designate an AOFF state, relying on the interchange between the clockwise (CW) and anticlockwise (ACW) propagation directions of the whispering gallery modes (WGMs) supported by the microdisk. Assuming that the microdisk laser works initially in the CW dominant state (Fig. 3d–I), the ACW mode is suppressed, and the optical power measured at the left side of the SOI bus waveguide is high. When an optical reset pulse is injected (Fig. 3d–II), it will invoke the ACW mode, which will be retained even after the reset pulse has passed through the microdisk laser, as shown in Fig. 3d–III. In this case, the power monitored at the left side of the SOI waveguide becomes low. Switching back to the CW dominant state can be achieved by injecting an optical set pulse from the right side of the SOI waveguide, as shown in Fig. 3d-IV,V. Microdisk-laser-based memories comprise a highly compact integrated memory scheme that has low switching power requirements (1.8 fJ) and fast switching times (~60 ps) but requires additional power for thermal tuning (~0.8 mW/bit) that increases the total energy consumption.

InP buried heterostructure (BH) photonic crystal (PhC) laser/ nanocavity

Significant research efforts have been invested in recent years in investigating (a) InP BH-PhC lasers targeting all-optical signal processing and next-generation optical packet switching systems and (b) nanocavities toward achieving successful optical memory operation for various types of optical processing, including network routing. In 2011, the first optically pumped PhC laser-based AOFF was introduced relying on a wavelength injection-locking technique in an InGaAsP/InP BH-PhC laser that exhibited fast switching times of 60 ps and switching powers in the range of ~20–70 μW. A significant step in the advancement of optical memory was performed in 2012, when a BH-PhC nanocavity again integrated in InGaAsP platform material was used to demonstrate optical memory bistability with a record-low static energy consumption on the order of 30 nW. Figure 3e shows a cross-sectional electron micrograph image of a fabricated sample, the respective hysteresis response when the laser wavelength was detuned by an offset *d* from its resonance, and the output power (P_{out}) versus the input power (P_{in}) for different wavelengths. The proposed BH-PhC nanocavity memory was tested with short pulses that can, in principle, lead to attractive memory speeds of 40 Gb/s; however, the switch-OFF time reported was on the order of 7 ns owing to the slow carrier relaxation time in the cavity. This technology was also the first to demonstrate high-integration-density memory setups exploiting wavelength-division-multiplexing and yielding a 128-bit storage capacity. Recently, an InP photonic crystal nanocavity with an embedded InGaAsP active region demonstrated an all-optical memory with only 2.3 nW operating power requirements and unlimited storage time. Among the most important advantages of the InP BH-PhC nanocavity-based memory technology are certainly (a) the ultra-low-energy consumption and (b) the proven capability to produce multi-bit photonic memory chips and high integration, with the main drawback thus far being the rather long switch-OFF time, which has most likely restricted their application to high-speed data traffic.

Hybrid InP-on-SOI photonic crystal (PhC) laser

Photonic crystals (PhCs) represent a disruptive solution toward low-power nanophotonic circuitry, with the heterogeneous integration of PhC lasers having also been successfully employed for optical memory operation42,43,78. The first InP-on-SOI PhC laser-based memory setup was demonstrated for the first time in 2013 using an optical pumping scheme and reporting on the >2 s storage capability. More recently, this laser structure was shown to perform successfully even with true pseudo-random bit sequence (PRBS) data patterns in both fundamental logic functionalities, i.e., gating and latching. Figure 3f shows the PhC nanocavity laser device and its principle of operation when relying on wavelength bistability through injection locking, depicting an indicative hysteresis loop formed for a given wavelength detuning. The device requires a constant optical bias signal and operates as a set-reset AOFF, taking advantage of the three discrete areas of injection signal optical power levels, as shown in Fig. 3f: Area I, where the injection power levels allow for the set operation, as the laser output is changed from a free-running (unlocked) to an injection-controlled (locked) state; Area II, where the injection power levels are below a certain threshold, enabling the reset operation, i.e., the laser output returns to its free-running (unlocked) state, and Area III, where the injection power levels cover the bistable range and enable the storing operation, because the laser emission retains its previous state. Hybrid InP-on-SOI PhCs combine some important advantages for memory applications as they can satisfy at the same time three critical requirements: (a) low footprint (6.4 μm^2), (b) low-energy consumption (13 fJ) and (c) high-speed bit-level operation, which have all been already verified experimentally at up to 10 Gb/s with the true data traffic. Considering that this memory technology can, in principle, be migrated to an electrically pumped scheme similar to the respective electrically pumped PhC laser nanocavities demonstrated more recently. This platform seems to hold all the necessary credentials toward promising optical memories for real application needs.

SUMMARY

- Optical memory is an electronic storage medium that uses a laser beam to store and retrieve the data. If we classify the memory system then optical memory comes under the external memory in the computer system.
- The optical memory is a round disk with a shiny surface on which data is imprinted by means of a laser beam. Thus optical technology was first used to represent analog sound signals into digital form.
- Computer memory operates at a high speed compared to storage that is slower but offers higher capacities. If needed, contents of the computer memory can be transferred to storage; a common way of doing this is through a memory management technique called virtual memory.
- The information density of a Compact Disc is relatively high, and is related to the wavelength of the laser, which is used, as well as other parameters. To allow as much information as possible to be stored, it is therefore important to use a laser with the shortest possible wavelength.
- This compound changes its physical state on heating depending on the temperature applied. When the temperature rises above its melting point (around 600°) it changes to liquid form and in its crystallization point (around 200°) it becomes solid.
- CD burning is an amazing technology and its working principle is really fascinating. CD burners encode music and other datas on to the blank CD which can be retrieved through CD reading.
- CDs are susceptible to damage from both daily use and environmental exposure. Pits are much closer to the label side of a disc, so that defects and dirt on the clear side can be out of focus during playback.
- The bottom layer of the compact disk that has exposure to the laser beam is formed by the transparent polycarbonate plastic. This appears as the clear glass base and its surface is programmed to store data.
- The laser beam with high intensity is focused on the disk which imprints or indents the series of microscopic *pits* on the surface of polycarbonate plastic. These pits are arranged on a long circular track on the surface of the disk spiraling from the center of the disk to the outer edge thus forming a master disk.

- The low power laser beam is focused on the polycarbonate plastic. The laser beam travels through the polycarbonate plastic layer and gets reflected back from the aluminium layer.
- The tracks are implemented on both sides of DVDs. This structure can be understood as the two single-sided disks that are put together to form a sandwich but the topmost disk is turned upside down.
- The high definition optical disk is the advanced version of DVDs that are used to store high definition videos. To raise the capacity the laser beam with an even shorter wavelength in the blue-violet range is used. Pits on high definition optical disk are even shorter when compared to DVD.
- Compact Disc Digital Audio (CDDA or CD-DA), also known as Digital Audio Compact Disc or simply as Audio CD, is the standard format for audio compact discs.

REFERENCES

1. Amy HarmonPublished: 12 October 2003 (12 October 2003). "MUSIC; What Price Music?". Nytimes.com. Archived from the original on 12 May 2013. Retrieved 25 July 2013.
2. Andy McFadden (8 August 2007). "CD-Recordable FAQ". Archived from the original on 20 September 2007. Retrieved 20 September 2007.
3. Bosch, Xavier (2001). "Fungus eats CD". Nature. Archived from the original on 12 December 2013.
4. Campaign For Digital Rights (5 December 2001). "Copy Protected CDs (via Archive.org)". Archived from the original on 5 December 2001.
5. K. Schouhamer Immink (2007). "Shannon, Beethoven, and the Compact Disc". IEEE Information Theory Society Newsletter. 57: 42–46. Retrieved 6 February 2018.
6. Sharpless, Graham (July 2003). "Introduction to CD and CD-ROM" (PDF). Deluxe Global Media Services Ltd. Archived (PDF) from the original on 9 March 2016. Retrieved 3 May 2016.
7. Tim Buthe and Walter Mattli, The New Global Rulers: The Privatization of Regulation in the World Economy, Princeton University Press, Feb. 2011.

CHAPTER 4

HOLOGRAPHICS

INTRODUCTION

Holography is the science of making holograms which are usually intended for displaying three dimensional images. It is a physical structure that diffracts light into an image. A holographic image can be seen looking into an illuminated holographic print by shining a laser through a hologram and projecting an image on the screen.

A hologram is made by superimposing a second wavefront (normally called the reference beam) on the wavefront of interest, thereby generating an interference pattern which is recorded on a physical medium. When only the second wavefront illuminates the interference pattern, it is diffracted to recreate the original wavefront. Holograms can also be computer-generated by modelling the two wavefronts and adding them together digitally. The resulting digital image is then printed onto a suitable mask or film and illuminated by a suitable source to reconstruct the wavefront of interest.

4.1 OVERVIEW AND HISTORY OF HOLOGRAPHICS

Holography, means of creating a unique photographic image without the use of a lens. The photographic recording of the image is called a hologram, which appears to be an unrecognizable pattern of stripes and whorls but which—when illuminated by coherent light, as by a laser beam—organizes the light into a three-dimensional representation of the original object.

An ordinary photographic image records the variations in intensity of light reflected from an object, producing dark areas where less light is reflected and light areas where more light is reflected. Holography, however, records not only the intensity of the light but also its phase, or the degree to which the wave fronts making up the reflected light are in step with each other, or coherent. Ordinary light is incoherent—that is, the phase relationships between the multitude of waves in a beam are completely random; wave fronts of ordinary light waves are not in step.

The Hungarian-British physicist Dennis Gabor (in Hungarian: *Gábor Dénes*) was awarded the Nobel Prize in Physics in 1971 "for his invention and development of the holographic method".

His work, done in the late 1940s, was built on pioneering work in the field of X-ray microscopy by other scientists including Mieczysław Wolfke in 1920 and William Lawrence Bragg in 1939. This discovery was an unexpected result of research into improving electron microscopes at the British Thomson-Houston Company (BTH) in Rugby, England, and the company filed a patent in December 1947 (patent GB685286). The technique as originally invented is still used in electron microscopy, where it is known as electron holography, but optical holography did not really advance until the development of the laser in 1960. The word *holography* comes from the Greek words ὅλος (*holos*; "whole") and γραφή (*graphē*; "writing" or "drawing").

A hologram is a recording of an interference pattern which can reproduce a 3D light field using diffraction. The reproduced light field can generate an image which still has the depth, parallax, and other properties of the original scene. A hologram is a photographic recording of a light field, rather than an image formed by a lens. The holographic medium, for example the object produced by a holographic process (which may be referred to as a hologram) is usually unintelligible when viewed under diffuse ambient light. It is an encoding of the light field as an interference pattern of variations in the opacity, density, or surface profile of the photographic medium. When suitably lit, the interference pattern diffracts the light into an accurate reproduction of the original light field, and the objects that were in it exhibit visual depth cues such as parallax and perspective that change realistically with the different angles of viewing. That is, the view of the image from different angles represents the subject viewed from similar angles. In this sense, holograms do not have just the illusion of depth but are truly three-dimensional images.

The development of the laser enabled the first practical optical holograms that recorded 3D objects to be made in 1962 by Yuri Denisyuk in the Soviet Union and by Emmett Leith and Juris Upatnieks at the University of Michigan, USA. Early holograms used silver halide photographic emulsions as the recording medium. They were not very efficient as the produced grating absorbed much of the incident light. Various methods of converting the variation in transmission to a variation in refractive index (known as «bleaching») were developed which enabled much more efficient holograms to be produced.

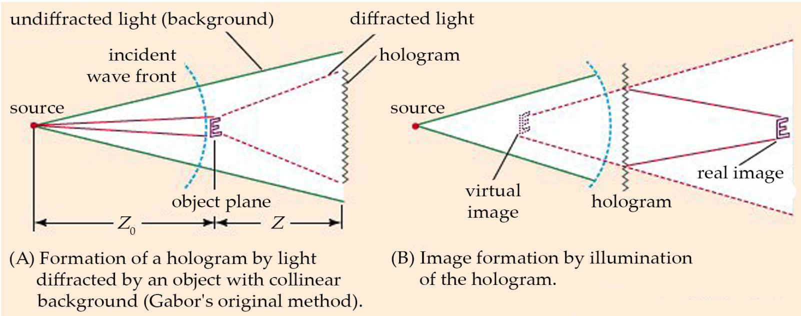


Figure 1. Gabor's original method for creating holograms.

Optical holography needs a laser light to record the light field. In its early days, holography required high-power and expensive lasers, but currently, mass-produced low-cost laser diodes, such as those found on DVD recorders and used in other common applications, can be used to make holograms and have made holography much more accessible to low-budget researchers, artists and dedicated hobbyists. A microscopic level of detail throughout the recorded scene can be reproduced. The 3d image can, however, be viewed with non-laser light. In common practice, however, major image quality compromises are made to remove the need for laser illumination to view the hologram, and in some cases, to make it. Holographic portraiture often resorts to a non-holographic intermediate imaging procedure, to avoid the dangerous high-powered pulsed lasers which would be needed to optically "freeze" moving subjects as perfectly as the extremely motion-intolerant holographic recording process requires. Holograms can now also be entirely computer-generated to show objects or scenes that never existed. Most holograms produced are of static objects but systems for displaying changing scenes on a holographic volumetric display are now being developed.

Holography is distinct from lenticular and other earlier autostereoscopic 3D display technologies, which can produce superficially similar results but are based on conventional lens imaging. Images requiring the aid of special glasses or other intermediate optics, stage illusions such as

Pepper›s Ghost and other unusual, baffling, or seemingly magical images are often incorrectly called holograms.

It is also distinct from specular holography which is a technique for making three-dimensional images by controlling the motion of specularities on a two-dimensional surface. It works by reflectively or refractively manipulating bundles of light rays, not by using interference and diffraction.

4.2 HOLOGRAM

The hologram, that is, the medium which contains all the information, is nothing more than a high contrast, very fine grain, black and white photographic film. Silver halide emulsion much like the black and white film you can buy in your neighborhood drug store. The film designed especially for holography is capable of very high resolution. One way of judging resolution of film or lenses is to see how many distinguishable lines can be resolved within a certain width, in this case it's a millimeter. A good film designed for holography is able to resolve up to 3000 lines/mm. The wavelengths of light from a He-Ne laser are approximately 24 micro-inches or twenty four millionths of an inch long, thus the need for such fine grain or high resolving power.

A hologram is an image recorded on a photosensitive plate or flat sensitive film. In a hologram a three dimensional (3D) image is recorded which is different from the ordinary or conventional photography.

A photograph is basically the recording of the differing intensities of the light reflected by the object and imaged by a lens. The light is incoherent, therefore, there are many different wavelengths of light reflecting from the object and even the light of the same wavelength is out of phase. There is a point to point correspondence between the object and the emulsion.

In the ordinary photography imaging technique the intensity distribution in the original scene is recorded. Therefore all the information about the optical paths to different parts of the scene is lost.

Any object to be recorded can be thought of as the sum of billions of points on the object which are reflecting more or less light. The lens of the camera focuses each object point to a corresponding point on the film and there it exposes a proportional amount of silver halide. Thus, your record is of the intensity differences on the object which form a pattern that one may ultimately recognize as the object photographed, in holography we are working with light waves and with, most likely, a silver halide film, yet, beyond that it is very difficult to compare the two. If we were to simply illuminate our object with laser light and take a photograph, we would still only be recording the different light intensities of the object; we would not have captured any information about the phase of the light waves after bouncing off the object.

In holography technique case

- Both the phase and the amplitude waves coming from the object are recorded.
- It uses coherent light illumination and using reference beam to convert the phase information into variation of intensity

We need a standard or reference. In the same way that a surveyor needs a reference point in order to make his measurements, we need a standard or a reference source in order to record the phase difference of the light waves and thus capture the information which supplies the vital dimensions and depth, to the holographic presentation. This standard we call a reference beam and it is supplied by the laser light itself. The reference light is emitted in what we call a plane wave. By enlisting the aid of a beam splitter we are able to form two beams. The reference beam is allowed to hit the film directly. It might be spread with a lens and aimed at the film by a mirror, but for all practical purposes this does not affect the light waves.

The other beam which we will refer to as the object or scene beam is also usually spread by a lens and guided by a mirror but it is directed at the object being holographed. As soon as object beam hits the object it is changed, or modulated according to the physical characteristics and dimensions of the object. So that the light which ultimately reaches the film plane after being reflected by the object now deviates in intensity and phase, from the virtually unhampered reference beam. That difference is a function of the object. What once began as a plane wave is now bouncing off the object in a complex wave front which consists of the summation of the multitude of infinitesimal object points reflecting light.

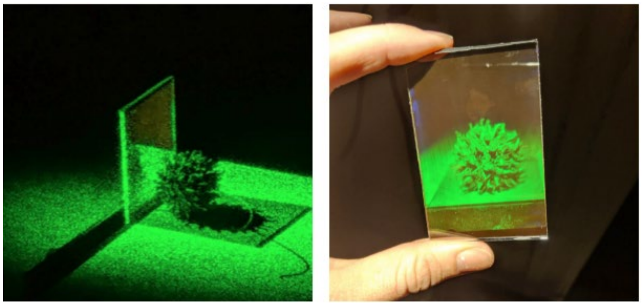
4.2.1 Types of Holograms

In the reflection hologram only selected wavelengths (colors) are reconstructed while the transmission hologram diffracts all the wavelengths of light so can it can have a rainbow appearance. The hologram does not change the wavelength (color) of light but controls how light is re-directed.

There are a number of distinct types of art holograms that can be defined by their optical-geometry and the recording medium.

The Denisyuk Reflection Hologram

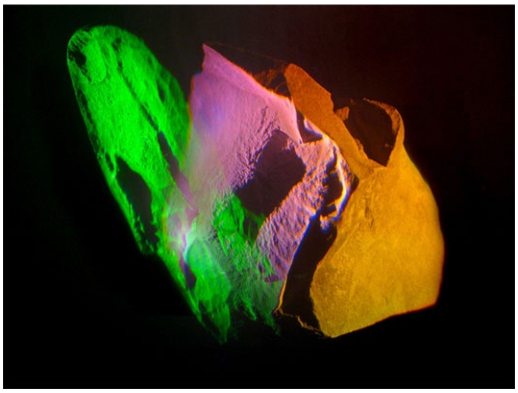
A *Denisyuk* hologram is created with a 'single' beam' of laser light that is shone through the hologram surface and then bounces off object back into the hologram surface.



Unless the physical structure is chemically manipulated the reflection hologram will appear the color that it is recorded. To create a true color refection hologram a multiple laser colors are combined – usually a red, green and blue.

Pseudo-color Reflection Hologram

Hologram are images of light, so colors are controlled by selecting and combining spectral colors. The holographic fringe structure can be expanded or shrunk changing the color of the reconstructed image. Pseudo-color reflection holograms can be created through multiple exposures between which the emulsion is swollen or shrunk to shift the recorded fringe spacing and therefore color,

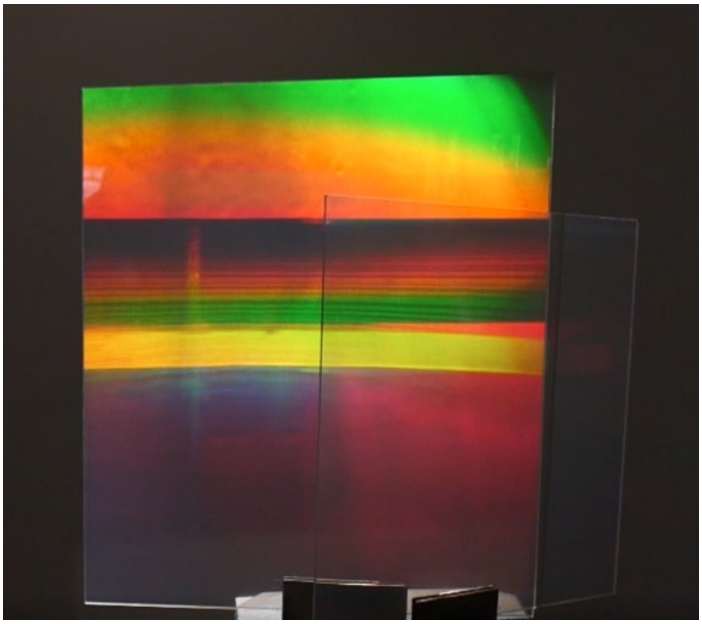


Laser Viewable Transmission Holograms

The laser viewable transmission hologram allows for a near perfect reconstruction of the optical field. This means that the recorded scene appears behind the film, and when replayed by a laser this scene can be very deep and sharp. These holograms are also used as a master recording that can then be transferred into a reflection or transmission holographic print.

Rainbow Holograms

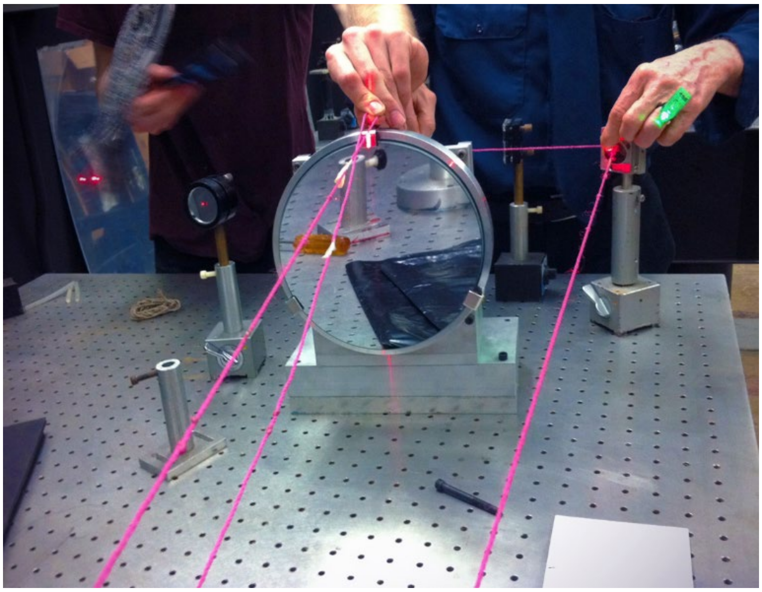
Transmission holograms have a different a visual quality and the color is controlled by geometry rather than chemistry. When illuminated with a white (broad spectrum) light source the transmission hologram will diffract all the wavelengths of light into the image. However as red-wavelengths are longer, they are deflected more than blue-wavelengths and so the image will have some color smear – the light spreads into a rainbow image.



A rainbow hologram, recorded from a horizontal master strip and replayed with a light from above, will have horizontal parallax – enabling the spatial qualities when looking with two eyes or moving side-to-side. But moving up-and-down produces no change in spatial perspective and viewer only sees a change of color. By recording a number of masters on an achromatic angle the spectral colors can be recombine to produce images that are whitish (achromatic) or color-mixed (ie, RGB red+green+blue).

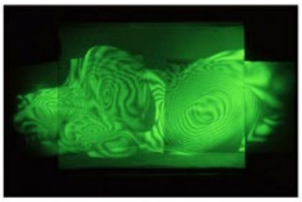
Pulse Laser Holography

Recording a hologram requires the interference pattern to be stationary during the exposure. For this reason holograms are traditionally made of static sculptures on vibration isolated tables. A pulse laser produces an ultra short flash of light, thus freezing motion and allowing for holograms of live subjects.



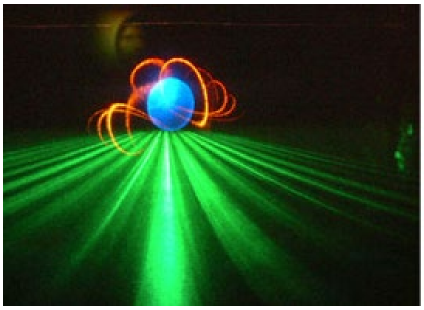
Holographic Interferometry

As the holographic recording process is dependent on the interference pattern between optical-waves the holographic image is the comparative shape between these waves. Holographic interferometry utilizes this property to detected small variations in form, with applications in industrial non-destructive testing. As part of her 'Strata Series' Sally Weber used a holographic interferometry technique – a double laser pulse – to show the movements of breath and blood under the skin.


Holographic Optical Elements

Holograms can also be used to direct light, which has commercial application in lighting design as well as being used for video projection screens and AR displays. This property has been incorporated into the process of making holograms by artists. Before the stenciling and digital printing techniques became widely practiced, Rudie Berkhout was creating spatially dynamic holograms using multiple holographic optical elements (HOEs) to shape light into dynamic abstract images. Berkhout's work explores the optical landscape in relation to a cosmic nature of the perceptual field: "I like the work to oscillate between landscape and abstract painting, challenging viewers and jolting their usual perception of the world."




Stenciling and Multiplex Holography

Multiplexing is a technique of recording multiple holographic exposures across the surface of the master hologram. By transferring a multiplexed master, a final print can be produced where the 'virtual windows' onto the scene is fragmented allowing for spatial animation, stereopsis and depth perception by parallax as the viewer moves around.

Dot-Matrix Holograms

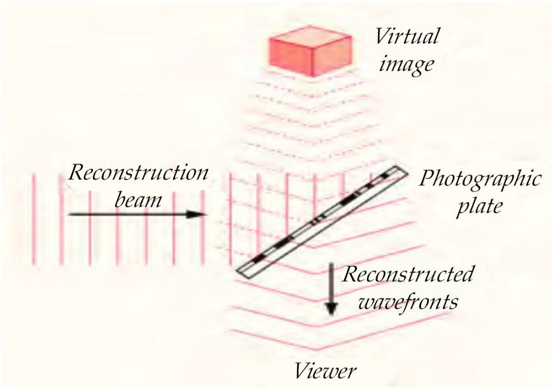
Dot-matrix hologram printing is a technique of building up an image of diffractive 'pixels'. Each area is recorded with a particular geometry that diffracts light by a corresponding angle. The illuminating light is deflected into varying divergent color-spectrums. The image is a tiling of colors which means the images are very bright but do not have 3-d depth information. Dot-matrix holograms have been used as a means of decorative 'light architecture'. Michael Bleyenberg defines 'light architecture': "This term indicates a vision: to 'plan and construct' environments beyond everyday perception and experience, barely tangible, not using solid material, but the ephemeral medium light."

Computer Generated 'Digital' Holography

With computer generated 'digital' holograms the fringe pattern of each pixel is calculated and recorded into the hologram. There are a number of ways of making digital holograms in which small regions of the film, termed 'Voxels' or 'Hogels', are exposed to the pre-calculated fringe pattern, such as by using a spatial light modulator (SLM) or electron beam lithography. In early systems these Hogels were noticeable causing the surface of the hologram to look pixelated.

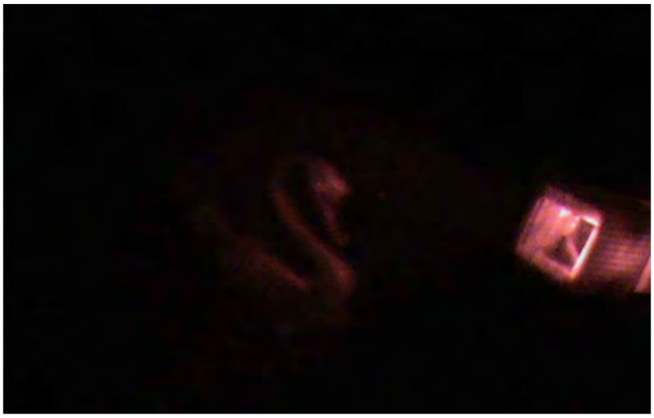
4.2.2 Reconstruction of Hologram

When the processed holographic film is illuminated once again with the reference beam, diffraction from the fringe pattern on the film reconstructs the original object beam in both intensity and phase. Because many viewpoints are stored, each of the viewer's eyes sees the image from a slightly different angle, so the image appears three-dimensional. This is known as stereopsis.



The viewer can move his or her viewpoint and see the image rotate exactly as the original object would. The central miracle of holography is that when the recorded grating is later illuminated by a substitute reference beam, the original object beam is reconstructed, producing a 3D image.

a) interference patterns



b) reconstruction of picture a)

Mass Replication

An existing hologram can be replicated, either in an optical way similar to holographic recording, or in the case of surface relief holograms, by Mass replication embossing. Surface relief holograms are recorded in photoresists or photothermoplastics, and allow cheap mass reproduction. Such embossed holograms are now widely used, for instance as security features on credit cards or quality merchandise.

The first step in the embossing process is to make a stamper by electrodeposition of nickel on the relief image recorded on the photoresist or photothermoplastic. When the nickel layer is thick enough, it is separated from the master hologram and mounted on a metal backing plate. The material used to make embossed copies consists of a polyester base film, a resin separation layer and a thermoplastic film constituting the holographic layer. The embossing process can be carried out with a simple heated press. The bottom layer of the duplicating film (the thermoplastic layer) is heated above its softening point and pressed against the stamper so that it takes up its shape. This shape is retained when the film is cooled and removed from the press. In order to permit the viewing of embossed holograms in reflection, an additional reflecting layer of aluminium is usually added on the hologram recording layer.

Dynamic Holography

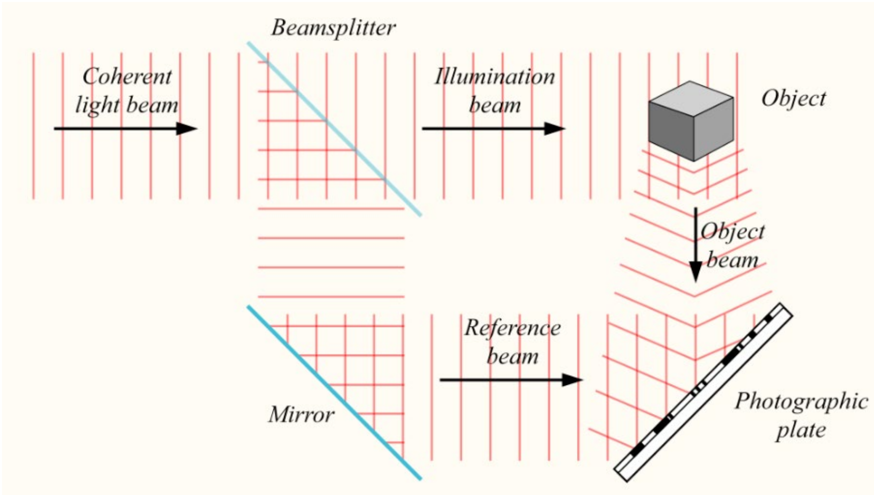
There exist also holographic materials which don't need the developing process and can record a hologram in a very short time. This allows to use holography to perform some simple operations in an all-optical way.

Examples of applications of such real-time holograms include phase-conjugate mirrors ("time-reversal" of light), optical cache memories, image processing (pattern recognition of time-varying images), and optical computing. The amount of processed information can be very high (terabit/s), since the operation is performed in parallel on a whole image. This compensates the fact that the recording time, which is in the order of a μs, is still very long compared to the processing time of an electronic computer. The optical processing performed by a dynamic hologram is also much less flexible than electronic processing. On one side one has to perform the operation always on the whole image, and on the other side the operation a hologram can perform is basically either a multiplication or a phase conjugation. But remember that in optics, addition and Fourier transform are already easily performed in linear materials, the second simply by a lens. This enables some applications like a device that compares images in an optical way. The search for novel nonlinear optical materials for dynamic holography is an active area of research. The most common materials are photorefractive crystals, but also in semiconductors or semiconductor heterostructures (such as quantum wells), atomic vapors and gases, plasmas and even liquids it was possible to generate holograms.

A particularly promising application is optical phase conjugation. It allows to remove the wavefront distortions a light beam receives when passing through an aberrating medium, by sending it back through the same aberrating medium with a conjugated phase. This is useful for example in free-space optical communications to compensate the atmospheric turbulence (the phenomenon that gives rise to the twinkling of starlight).

4.3 WORKING PRINCIPLE OF HOLOGRAPHY

Holography is a technique that enables a light field (which is generally the result of a light source scattered off objects) to be recorded and later reconstructed when the original light field is no longer present, due to the absence of the original objects. Holography can be thought of as somewhat similar to sound recording, whereby a sound field created by vibrating matter like musical instruments or vocal cords, is encoded in such a way that it can be reproduced later, without the presence of the original vibrating matter. However, it is even more similar to Ambisonic sound recording in which any listening angle of a sound field can be reproduced in the reproduction.



4.3.1 Laser

In laser holography, the hologram is recorded using a source of laser light, which is very pure in its color and orderly in its composition. Various setups may be used, and several types of holograms can be made, but all involve the interaction of light coming from different directions and producing a microscopic interference pattern which a plate, film, or other medium photographically records.

In one common arrangement, the laser beam is split into two, one known as the object beam and the other as the reference beam. The object beam is expanded by passing it through a lens and used to illuminate the subject. The recording medium is located where this light, after being reflected or scattered by the subject, will strike it. The edges of the medium will ultimately serve as a window through which the subject is seen, so its location is chosen with that in mind. The reference beam is expanded and made to shine directly on the medium, where it interacts with the light coming from the subject to create the desired interference pattern.

Like conventional photography, holography requires an appropriate exposure time to correctly affect the recording medium. Unlike conventional photography, during the exposure the light source, the optical elements, the recording medium, and the subject must all remain motionless relative to each other, to within about a quarter of the wavelength of the light, or the interference pattern will be blurred and the hologram spoiled. With living subjects and some unstable materials, that is only possible if a very intense and extremely brief pulse of laser light is used, a hazardous procedure which is rare and rarely done outside of scientific and industrial laboratory settings. Exposures lasting several seconds to several minutes, using a much lower-powered continuously operating laser, are typical.

4.3.2 Apparatus

A hologram can be made by shining part of the light beam directly into the recording medium, and the other part onto the object in such a way that some of the scattered light falls onto the recording medium. A more flexible arrangement for recording a hologram requires the laser beam to be aimed through a series of elements that change it in different ways. The first element is a beam splitter that divides the beam into two identical beams, each aimed in different directions:

- One beam (known as the 'illumination' or 'object beam') is spread using lenses and directed onto the scene using mirrors. Some of the light scattered (reflected) from the scene then falls onto the recording medium.
- The second beam (known as the 'reference beam') is also spread through the use of lenses, but is directed so that it does not come in contact with the scene, and instead travels directly onto the recording medium.

Several different materials can be used as the recording medium. One of the most common is a film very similar to photographic film (silver halide photographic emulsion), but with a much higher concentration of light-reactive grains, making it capable of the much higher resolution that holograms require. A layer of this recording medium (e.g., silver halide) is attached to a transparent substrate, which is commonly glass, but may also be plastic.

4.3.3 Process

When the two laser beams reach the recording medium, their light waves intersect and interfere with each other. It is this interference pattern that is imprinted on the recording medium. The pattern itself is seemingly random, as it represents the way in which the scene's light *interfered* with the original light source – but not the original light source itself. The interference pattern can be considered an encoded version of the scene, requiring a particular key – the original light source – in order to view its contents.

This missing key is provided later by shining a laser, identical to the one used to record the hologram, onto the developed film. When this beam illuminates the hologram, it is diffracted by the hologram's surface pattern. This produces a light field identical to the one originally produced by the scene and scattered onto the hologram.

4.3.4 Comparison with Photography

Holography may be better understood via an examination of its differences from ordinary photography:

- A hologram represents a recording of information regarding the light that came from the original scene as scattered in a range of directions rather than from only one direction, as in a photograph. This allows the scene to be viewed from a range of different angles, as if it were still present.
- A photograph can be recorded using normal light sources (sunlight or electric lighting) whereas a laser is required to record a hologram.
- A lens is required in photography to record the image, whereas in holography, the light from the object is scattered directly onto the recording medium.
- A holographic recording requires a second light beam (the reference beam) to be directed onto the recording medium.
- A photograph can be viewed in a wide range of lighting conditions, whereas holograms can only be viewed with very specific forms of illumination.
- When a photograph is cut in half, each piece shows half of the scene. When a hologram is cut in half, the whole scene can still be seen in each piece. This is because, whereas each point in a photograph only represents light scattered from a single point in the scene, *each point* on a holographic recording includes information about light scattered from *every point* in the scene. It can be thought of as viewing a street outside a house through a 120 cm × 120 cm (4 ft × 4 ft) window, then through a 60 cm × 120 cm (2 ft × 4 ft) window. One can see all of the same things through the smaller window (by moving the head to change the viewing angle), but the viewer can see more *at once* through the 120 cm (4 ft) window.
- A photograph is a two-dimensional representation that can only reproduce a rudimentary three-dimensional effect, whereas the

reproduced viewing range of a hologram adds many more depth perception cues that were present in the original scene. These cues are recognized by the human brain and translated into the same perception of a three-dimensional image as when the original scene might have been viewed.

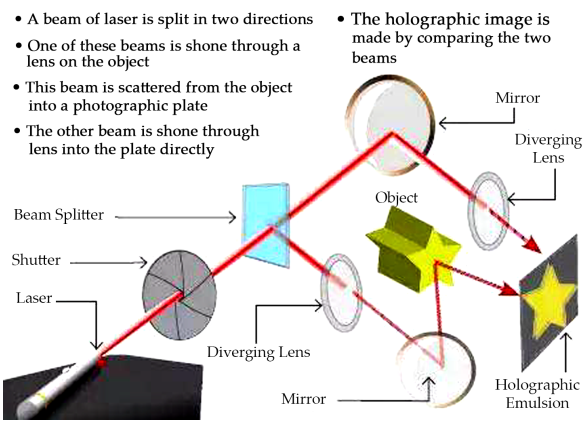
- A photograph clearly maps out the light field of the original scene. The developed hologram's surface consists of a very fine, seemingly random pattern, which appears to bear no relationship to the scene it recorded.

4.3.5 Advancements in Holographic Display Technology

A hologram uses light diffraction to create an image. In normal photography, a lens is used to focus on an image, and then light, dark, and colored areas of the image (seen as reflected light) are recorded on film or a digital sensor. By contrast, holography is a photographic technique that records the shape of light waves as they bounce off an object.

When light waves meet, they interfere (in the same way ripples of water interfere with each other if you throw two pebbles into a pond). This light wave interference pattern (a light field) is then recorded to make a hologram. Because lasers create pure, coherent light, they make it possible to accurately record the light wave interference patterns and recreate a 3D image from them.

Essentially, a hologram is created when a beam of laser light is split in two, with one of the resulting beams shone on an object (the object beam) then scattered onto a photographic plate, while the other beam is directed directly onto the plate (a reference beam). A holographic image is derived from the differences in the two beams.

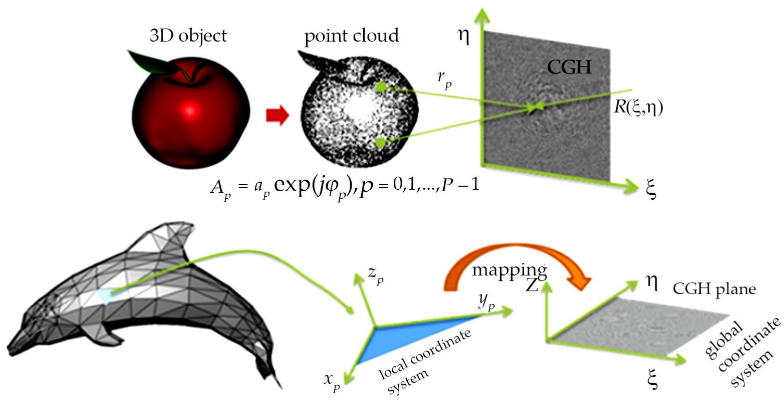


The Challenge of Three Dimensions

However, while standard holographic images are three-dimensional in appearance, they are still essentially two-dimensional because they are viewed on the flat screen of a display monitor or projected onto a flat surface such as a wall or sheet of glass. A viewer must look directly (or within a limited view angle range) at the flat surface to view the projection. While the image appears three-dimensional, it is technically a 3D projection, not a hologram.

It turns out that creating 3D holograms "in which a 3-D scene is encoded in terms of optical diffraction, transformed into the fringe patters of the hologram that is further converted into a signal for a spatial light modulator (SLM) and displayed in real time, is an extremely challenging enterprise."[2] For the last 40 years, developers have been inspired by Princess Leia to try and create true 3D holograms that can be projected into thin air—also called "volumetric images"—which would be viewable from all sides with 360-degree visual dimensionality.

In addition to the optical challenges of capturing and projecting 3D holograms taken from real-world images, the industry is also grappling with the challenges of creating computer-generated content that can be viewed in 3D holographic format. Mathematical models of 3D objects such as point cloud algorithms and polygon-based algorithms can be used, but they create a large computational load for any holographic projection devices to process in real time.



Holographic Projections and Displays

Traditional projectors work by passing light through a graphical image that blocks some of the light, creating shading and color of the projected picture. Holographic projectors create a picture via refraction as light passes through the recorded interference pattern. Light shining through a flat interference pattern (recorded on a holographic plate as shown above) produces an image that has three-dimensional qualities but is still flat.

To project a true 3D image, a spinning mirror system can be used. First developed by researchers at University of Southern California's Institute for Creative Technology, the technique uses a high-speed video projector aimed towards a spinning platform with a mirror attached, a holographic diffuser, and hardware for decoding digital images.4 Because the mirror spins rapidly, it reflects in all directions, allowing viewers at any angle to see the image in three dimensions.

A holographic interference pattern works only with one color, since the interference pattern is recorded using a single wavelength of light. To create colored holograms, multiple interference patterns must be recorded using different wavelengths, and then holographic projectors use colored lasers that illuminate the corresponding interference patterns for their respective colors.

Holographic images can be projected onto glass or plexiglass, or more unconventional "screens" such as a semi-transparent net or even a cloud of smoke.[5] Holograms that appear within videos and movies, or in gaming and AR/VR applications are, of course, shown on whatever device or display screen is being used for that content, whether a television set or a pair of smart glasses. Whatever the projection of display medium, 3D images shown can be captured using holographic photography approaches or may be entirely computer-generated content.

The State of 3D Holograms Today

People across industries are excited by the potential of 3D holographic technology, including television, gaming, medical imaging, computer-aided design, automated robot systems, air traffic control, education, and cultural heritage dissemination. The potential for 3D interactive visuals and remote communication is enticing.

Many "holograms" currently on the market are still technically 3D images shown on a 2D surface, real as they may appear. For example, a hologram in a virtual reality application is still a 2D element, since it's being viewed on the flat display screen of a VR headset, even if it appears to be (and can be interacted with) in the viewer's perceived 3D environment. But full 3D holograms are emerging. Holographic science has advanced to the point of providing full-motion parallax (both vertical and horizontal), focus at different depths (providing all depth cues) that is viewable by multiple viewers with unique perspectives and wide viewing angles.

Creatively speaking, regardless of the technology used, many of the applications of holographic images we're seeing today are pretty exciting and realistic. There are many around already—from the holographic marker on credit cards and drivers' licenses, to authentication stickers on product packaging, to the latest prototype OLED transparent television screens.

Meanwhile, a proliferation of new technologies is being explored as ways to create volumetric 3D images. For example:

- Physicist Daniel Smalley at Brigham Young University has developed a method he describes as less like a hologram and more "like a high-speed Etch a Sketch". It uses the force of invisible laser beams to manipulate a particle of cellulose that creates images by moving at ultra-high speed. However, the technique is not yet safe enough for consumer use.
- South Korean scientists have found a way to produce 3D holographic displays by placing a titanium film covered with tiny pinholes behind an off-the-shelf LCD display screen.
- Japanese firm Aerial Burton created 3D touch-sensitive lighted-dot patterns using a plasma technology.

4.4 APPLICATIONS OF HOLOGRAPHY

Holography is widely used in various fields of application, including art, data storage, security and so on.

4.4.1 Art

Early on, artists saw the potential of holography as a medium and gained access to science laboratories to create their work. Holographic art is often the result of collaborations between scientists and artists, although some holographers would regard themselves as both an artist and a scientist.

Salvador Dalí claimed to have been the first to employ holography artistically. He was certainly the first and best-known surrealist to do so, but the 1972 New York exhibit of Dalí holograms had been preceded by the holographic art exhibition that was held at the Cranbrook Academy of Art in Michigan in 1968 and by the one at the Finch College gallery in New York in 1970, which attracted national media attention. In Great Britain, Margaret Benyon began using holography as an artistic medium in the late 1960s and had a solo exhibition at the University of Nottingham art gallery in 1969.This was followed in 1970 by a solo show at the Lisson Gallery in London, which was billed as the "first London expo of holograms and stereoscopic paintings".

During the 1970s, a number of art studios and schools were established, each with their particular approach to holography. Notably, there was the San Francisco School of Holography established by Lloyd Cross, The Museum of Holography in New York founded by Rosemary (Posy) H. Jackson, the Royal College of Art in London and the Lake Forest College Symposiums organised by Tung Jeong. None of these studios still exist; however, there is the Center for the Holographic Arts in New York and the HOLOcenter in Seoul, which offers artists a place to create and exhibit work.

During the 1980s, many artists who worked with holography helped the diffusion of this so-called "new medium" in the art world, such as Harriet Casdin-Silver of the United States, Dieter Jung of Germany, and Moysés Baumstein of Brazil, each one searching for a proper «language» to use with the three-dimensional work, avoiding the simple holographic reproduction of a sculpture or object. For instance, in Brazil, many concrete poets (Augusto de Campos, Décio Pignatari, Julio Plaza and José Wagner Garcia, associated with Moysés Baumstein) found in holography a way to express themselves and to renew Concrete Poetry.

A small but active group of artists still integrate holographic elements into their work. Some are associated with novel holographic techniques; for example, artist Matt Brand employed computational mirror design to eliminate image distortion from specular holography.

4.4.2 Data Storage

One of the application of holography is information storage. A single image can contain huge information which can be accessed by reflecting light from different angles unlike traditional optical storage methods like CD-ROM. Holograms are being extensively researched for high capacity data storage devices for entertainment and computing purposes. Though there are technical and material challenges to holographic data storage, they could yield capacities of a trillion bits per square centimeter. Fast progress in multiple-frequency holography offers atleast the theoritical potential for managing high data densities and readout speeds.

In holographic data storage, data is first turned into a two dimensional pattern of light by a spatial light modulator which is an array of light switches that can store as many as one million bits or pixels. Laser light is beamed through the SLM to the recording medium, such as a photopolymer, while a reference beam also illuminates the medium so that interference patterns are created. These expose the medium by generating corresponding differences in optical properties such as refractive index or absorption. Many pages of holograms can be multiplexed onto the same medium, either by varying the angle or phase of the reference beam, or by using different frequencies, among other techniques.

4.4.3 Dynamic Holography

In static holography, recording, developing and reconstructing occur sequentially, and a permanent hologram is produced.

There also exist holographic materials that do not need the developing process and can record a hologram in a very short time. This allows one to use holography to perform some simple operations in an all-optical way. Examples of applications of such real-time holograms include phase-conjugate mirrors ("time-reversal" of light), optical cache memories, image processing (pattern recognition of time-varying images), and optical computing.

The amount of processed information can be very high (terabits/s), since the operation is performed in parallel on a whole image. This compensates for the fact that the recording time, which is in the order of a microsecond, is still very long compared to the processing time of an electronic computer. The optical processing performed by a dynamic hologram is also much less flexible than electronic processing. On one side, one has to perform the operation always on the whole image, and on the other side, the operation a hologram can perform is basically either a multiplication or a phase conjugation. In optics, addition and Fourier transform are already easily performed in linear materials, the latter simply by a lens. This enables some applications, such as a device that compares images in an optical way.

The search for novel nonlinear optical materials for dynamic holography is an active area of research. The most common materials are photorefractive crystals, but in semiconductors or semiconductor heterostructures (such as quantum wells), atomic vapors and gases, plasmas and even liquids, it was possible to generate holograms.

A particularly promising application is optical phase conjugation. It allows the removal of the wavefront distortions a light beam receives when passing through an aberrating medium, by sending it back through the same aberrating medium with a conjugated phase. This is useful, for example, in free-space optical communications to compensate for atmospheric turbulence (the phenomenon that gives rise to the twinkling of starlight).

4.4.4 Security

Holography can also be used in security applications. Complex holograms are much more difficult to reproduce compared to static images. For this reason, holograms are often placed on valuable items to reduce the chances of forgery. Common security applications of holograph include labels on credit cards and images embedded in government currency. Important documents such as passports and identification cards also often use them to discourage illegal reproduction.

4.4.5 Holographic Sensors

A holographic sensor is a device that comprises a hologram embedded in a smart material that detects certain molecules or metabolites. This detection is usually a chemical interaction that is transduced as a change in one of the properties of the holographic reflection (as in the Bragg reflector), either refractive index or spacing between the holographic fringes. The specificity of the sensor can be controlled by adding molecules in the polymer film that selectively interacts with the molecules of interest.

4.4.6 Holographic interferometry

Holographic interferometry (HI) is a technique that enables static and dynamic displacements of objects with optically rough surfaces to be measured to optical interferometric precision (i.e. to fractions of a wavelength of light). It can also be used to detect optical-path-length variations in transparent media, which enables, for example, fluid flow to be visualized and analyzed. It can also be used to generate contours representing the form of the surface or the isodose regions in radiation dosimetry.[

It has been widely used to measure stress, strain, and vibration in engineering structures.

4.4.7 Interferometric Microscopy

The hologram keeps the information on the amplitude and phase of the field. Several holograms may keep information about the same distribution of light, emitted to various directions. The numerical analysis of such holograms allows one to emulate large numerical aperture, which, in turn, enables enhancement of the resolution of optical microscopy. The corresponding technique is called interferometric microscopy. Recent achievements of interferometric microscopy allow one to approach the quarter-wavelength limit of resolution.

4.4.8 Hologram Making Recent News

Recently thousands of people protested against a draconian law which endangered civil liberties in Spain. To challenge the legislation passed, the activists called on people all over the world to volunteer to become a hologram in the protest. It has been world's first holographic protest.

4.4.9 Other Applications

Holographic scanners are in use in post offices, larger shipping firms, and automated conveyor systems to determine the three-dimensional size of a package. They are often used in tandem with checkweighers to allow automated pre-packing of given volumes, such as a truck or pallet for bulk shipment of goods. Holograms produced in elastomers can be used as stress-strain reporters due to its elasticity and compressibility, the pressure and force applied are correlated to the reflected wavelength, therefore its color. Holography technique can also be effectively used for radiation dosimetry.

High Security Registration Plates

High-security holograms can be used on license plates for vehicles such as cars and motorcycles. As of April 2019, holographic license plates are required on vehicles in parts of India to aid in identification and security, especially in cases of car theft. Such number plates hold electronic data of vehicles, and have a unique ID number and a sticker to indicate authenticity.

4.5 HOLOGRAPHIC DATA STORAGE

To use volume holography as a storage technology, the digital data to be stored must be imprinted onto the object beam for recording, then retrieved from the reconstructed object beam during readout. The input device for the system is called a spatial light modulator, or SLM. The SLM is a planar array of thousands of pixels; each pixel is an independent optical switch that can be set to either block or pass light. The output device is a similar array of detector pixels, such as a charge-coupled device (CCD) camera or CMOS pixel array. The object beam often passes through a set of lenses that image the SLM pixel array onto the output pixel array, as Figure 2 shows.

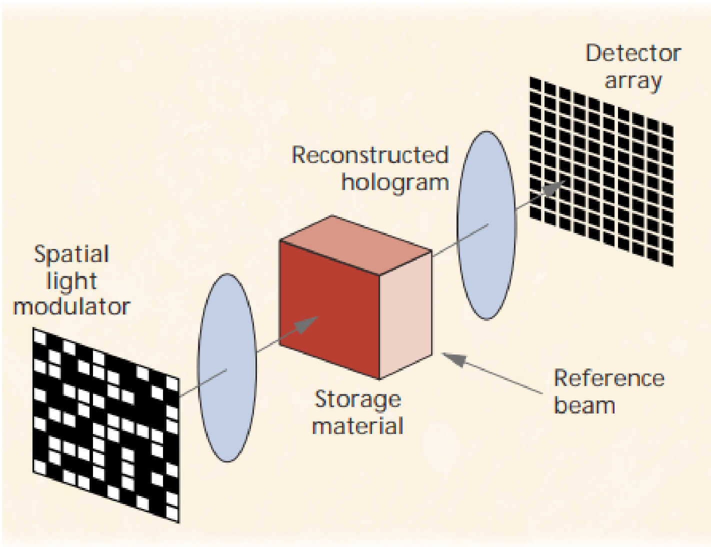


Figure 2. For data storage, information is put onto the object beam with a spatial light modulator and removed from a reconstructed object beam with a detector array.

The hologram can be formed anywhere in the imaging path between the input pixel array and the output pixel array. To maximize storage density, the hologram is usually recorded where the object beam attains a tight focus. When the reference beam reconstructs the hologram, the object beam continues along the original imaging path to the camera, where the optical output can be detected in parallel and converted to digital data. Capacity and readout rate are maximized when each detector pixel is matched to a single pixel on the SLM, but for large pixel arrays this requires careful optical design and alignment. Recently, the IBM Almaden Research Center built two test platforms that implement this pixel-to-pixel matching for input data masks of 1,024 × 1,024 pixels, as well as for smaller real-time SLMs.

4.5.1 Storage Materials

Photosensitive materials for volume holography are generally classified as either read-write or write-once.

Read-Write Materials

Most holographic read-write materials are inorganic photorefractive crystals doped with transition metals such as iron or rare-earth ions such as praseodymium, grown in large cylinders in the same way as semiconductor materials. Large samples can be cut and polished, making thick holograms possible. These materials react to the light and dark regions of an interference pattern by transporting and trapping photo-ionized electrons.

Through the linear electro-optic effect exhibited by these crystals, the electrical fields created by the trapped charge give rise to an index or phase grating suitable for diffracting light. Thus, the spatial variations in light intensity present in the interference pattern become identical variations in the index of refraction. The trapped charge can be rearranged by subsequent illumination, which makes it possible to erase recorded holograms and replace them with new ones. However, the ease of charge re-excitation also results in the gradual erasure of stored holograms during normal readout. In the dark, the lifetime of these holograms ranges from months to years as the trapped charge slowly leaks away.

Recorded holograms can be "fixed" (made semipermanent and resistant to erasure during readout) through thermal or electronic processes. The fixing process affects all the stored holograms within a volume simultaneously. Thus, individual pages of data cannot be erased and replaced this way. An alternative for achieving nonvolatile storage in photorefractive materials is to record at a light wave length not normally absorbed by the crystal except in the presence of a third "gating" beam of different wavelength. This beam is present only during the recording and is switched off for readout.

Organic photorefractive polymers have also been developed. These materials provide more opportunity for performance tuning because you can fabricate them using a wide variety of constituents. However, these materials tend to be limited in thickness and require large applied voltages.

Write-once materials

Writing permanent volume holograms generally involves irreversible photochemical reactions, triggered by the bright regions of the optical interference pattern. For example, a photopolymer material will polymerize (bind short monomer chains together to form long molecular chains) in response to optical illumination. In contrast, the molecules in a photochromic material undergo a change in their absorption behavior. Such materials are inexpensive to make in quantity. However, both types can have problems reproducing the object beam faithfully—the photopolymer because of shrinkage, the photochromic because of oversensitivity to average local intensity.

Careful system design can minimize these problems. One advantage of a photopolymer is that after recording, any leftover monomers can be disposed of without affecting the recorded holograms. A photochromic material, however, requires a separate chemical or optical step to disable the unused absorbing molecules after the holograms are recorded. Currently available versions of these write-once materials are thin (approximately 100 μm)—the difficulties in making thick samples include insufficient optical quality or excessive absorption. As we will show later, however, new multiplexing techniques for thin materials have made write-once photopolymers one of the leading candidates for the first holographic memory products.

Dynamic Range

In the readout process, the reconstructed hologram is imaged onto the output detector array, where the digital data is extracted from the detected signal. Noise can cause errors to occur during the detection process in various ways. The basic trade-off in volume holography is caused by the fixed noise floor and the finite dynamic range of the recording material. In other words,

- The electronic detection process at the camera contributes the same amount of noise, no matter how bright the hologram, and
- As the number of holograms or the readout rate increases, the amount of power diffracted into each hologram reconstruction decreases.

Even if all other noise sources are negligible, there will be a certain hologram strength at which the signal-to-noise ratio is inadequate for error-free detection. The number of detected electrons per pixel can be written as

$$\eta_{\text{electrons}} \;\alpha\; M/\#^2 P_{\text{readout}} \frac{t_{\text{readout}}}{M^2 N_{\text{pixels}}},$$

where M is the number of multiplexed holograms, N_{pixels} the number of pixels per hologram, $t_{readout}$ the integration time of the camera, $P_{readout}$ the power in the readout beam, and M/# a material/system constant, which measures dynamic range. The storage capacity is MN_{pixel} and the readout rate is N_{pixel} /$t_{readout}$. An increase in either of these parameters leads to a decrease in the number of signal electrons.10 Given the minimum acceptable number of signal electrons per pixel, we can maximize the capacity and readout rate by increasing $P_{readout}$ or M/#.

Different processes determine the M/# constant in photorefractives and write-once media. In a photorefractive crystal, the holograms' recording exposures must be carefully scheduled to record equal-strength holograms. The first hologram is made quite strong. This first hologram erases slowly while the other holograms are stored, and finishes at the same strength as the weakly written final hologram. Alternatively, all the holograms can be cycled several times. The equalized diffraction efficiency falls as one over the square of the number of holograms, with the M/# as the proportionality coefficient:

$$n = \left(\frac{M/\#}{M}\right)^2$$

The M/# constant in a photorefractive material becomes large if the holograms can be recorded faster than they erase. In iron-doped lithium niobate, a typical M/# might be 1. This implies that to store 1,000 holograms with 1 million pixels and read each in 1 millisecond, we need about 1W in the reference beam.

A write-once material has much in common with photographic film: After a finite amount of input energy, the material is completely exposed. Each hologram gets its share of the dynamic range as it is recorded, preserving the bright and dark regions of the interference fringes. For instance, in a photopolymer material, the photosensitivity saturates as the available supply of monomers is exhausted. It turns out that the diffraction efficiency of individual holograms, when M of them are multiplexed in a saturable medium such as a photopolymer, also follows the $(M/\#/M)^2$ relationship. The most commonly used polymer is DuPont's HRF-150. The 100-micron-thick version has a M/# of 6.5, which reduces the required readout power by a factor of 40.

Understanding Noise

In addition to detector noise, other factors can cause errors:

- The readout conditions change. This can occur, for instance, when the recording alters the recording material properties. This causes unwanted changes in the reference beam path between the time the hologram is recorded and the time it is reconstructed. Often, the reference beam angle or wavelength can be tuned to optimize the diffraction efficiency and partially compensate for this effect.
- The detector array doesn't line up with the pixel array in the reconstructed hologram. This includes errors in camera registration, rotation, focus, tilt, and image magnification.
- The detector is receiving undesired light, either from light scattering off the storage material, crosstalk from other stored holograms, or crosstalk between neighboring pixels of the same hologram. Although crosstalk contributions scale with the strength of the holograms, scattering depends only on readout power and the components' optical quality.
- There are brightness variations across the detected image. This can be a problem if a single threshold is used across the image to separate the pixels into bright and dark and to assign binary values. These fluctuations can be caused by the SLM, the optical imaging, or the original laser beams.

4.5.2 Combating Errors

Commercial storage products have user error rates as low as 10^{-15}. For a 1-Gbyte hard drive, a user might expect to read the entire drive 100,000 times before a single bit error. On the other hand, at the hardware level, errors occur at a rate of perhaps 1 in 10,000 bits read. The designers decrease the error rate that the user sees by storing redundant bits along with user data. The sequence of user and redundant bits forms an error correction code. ECC algorithms performed in hardware after the read head can detect if a few bits within each code word are in error and then pass on corrected user bits.

The redundant bits cause a slight sacrifice in the capacity of each individual hologram. This sacrifice is measured by the ECC code rate—the fraction of bits stored that are actual user data. However, the increase in the raw bit error rate (BER) that can be tolerated allows additional pages to be stored, which increases overall capacity. A typical ECC code with a code rate of 0.9 can handle an input data stream with a raw BER of 10^{-4}. It can output the user's data, stripped of redundant bits, with the desired BER of 10^{-15}. Since the ECC code rate drops quickly as the expected raw BER climbs above 10^{-3}, most holographic memory designs aim for a raw BER of 10–4 or so.

Keeping the raw BER at this target value takes careful engineering of the optical system combined with signal processing and modulation coding. Careful engineering of the optics alone is not generally the most cost-effective solution to reducing the BER to the target value.

Signal Processing

Signal processing works by considering the storage device as an imperfect transmission "channel" for data that tends to smear together the signal energy from multiple bits of user data. Knowledge of how this intermixing occurs can be applied at the output end to eliminate the crosstalk and reproduce the originally transmitted bit sequence. In a telecommunications application or a bit-serial storage device like a hard drive or DVD disk, the smearing takes place between signals adjacent in time. In holographic storage, the smearing occurs spatially in two dimensions, as light intended for a particular CCD pixel diffracts into neighboring pixels.

Signal processing techniques for holographic storage are therefore 2D extensions of the 1D techniques developed for bit-serial devices. Examples of signal processing techniques used in holographic memories are adaptive thresholding and normalization, equalization, filtering, and partial response precoding at the input.

Modulation Codes

A modulation code dictates the way in which bits of information are encoded into the channel as data signals. They are selected to facilitate the detection process and hence improve overall performance. For instance, in bit-serial devices, modulation codes are used to set upper and lower bounds on the frequency at which the signal level changes. In holographic storage, modulation codes are used to avoid pixel combinations that are prone to distortion and to create easy-to-detect pixel patterns.

A convenient encoding that facilitates detection is the organization of binary data into small blocks of pixels, such that the number of bright pixels is constant (usually half the pixels). The simplest example is differential encoding, in which 2 pixels convey 1 bit of information. Several modulation codes with higher code rate and performance, and thus higher complexity, have since been developed for holographic storage. These codes have been used to demonstrate as many as 1,200 superimposed holograms in lithium niobate ($LiNbO_3$) at a raw BER of 10^{-8}.

4.5.3 System Configurations

The storage capacity for the simple, angle-multiplexed memory we have been discussing is simply the product of the number of holograms superimposed times the number of pixels in each page. The number of pixels per page is currently limited to roughly one million. The dynamic range (M/#) of available materials and the desire for a reasonable readout rate limit the number of holograms to about 1,000. Therefore, the capacity of a single angle-multiplexed holographic memory module is only 1 Gbit.

We must increase the capacity to at least 1 terabit to have a system that is competitive with alternative technologies. We can accomplish this by constructing a large memory consisting of multiple 1-Gbit modules. This technique is called spatial multiplexing, because multiple "stacks" of holograms are stored in different spatial locations of the recording material. Spatial multiplexing has several configuration options: holographic random-access memory (HRAM), compact modular holographic memory, and holographic 3D disks.

Holographic random-access memory

One approach for spatial multiplexing steers the reference and object beams to a stationary block of material containing multiple storage locations, as Figure 3 shows. With nonmechanical optical scanners, the HRAM system can very rapidly steer the optical beams. Most nonmechanical beam steerers use either an acousto-optic deflector or a one-dimensional liquid crystal SLM. By using large lenses (not shown in Figure 3), the information stored at separate locations can be directed back to a single detector array.

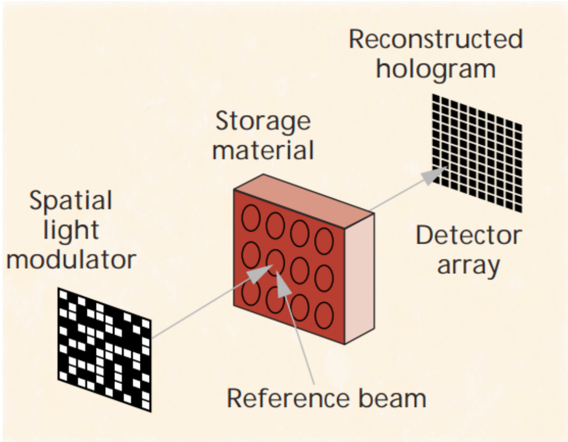


Figure 3. An HRAM system increases capacity by steering the object and reference beams to one of many storage locations on a large block of holographic recording material.

To construct a Tbit memory using this approach, we need 1,000 spatial locations (arranged in 2D as a 33 × 33 array), with each location storing 1 Gbit. The main challenge in building such a system is the optics that have to simultaneously transfer data from each of the 1,000 recording sites on the recording material to a single detector array. This will require considerable engineering improvements over present systems.

Recording Rate

The photosensitivity of most photorefractive crystals is relatively low; therefore, the recording rate is invariably one to two orders of magnitude slower than the readout rate. In addition, it is practically impossible to change the state of a single pixel within a stored hologram, and it is possible but not easy to replace a single hologram within a hologram stack.4 Instead, an entire stack of holograms must be erased together, either by heating or by illumination with the "gating" light. Therefore, an HRAM system is not truly a read-write memory; more accurately, it is an erasable write-once, read-many memory.

Readout Rate

The readout rate in an HRAM system is mostly limited by the camera integration time. The reference beam reconstructs the same hologram until a sufficient number of photons accumulate to differentiate bright and dark pixels. An oft-mentioned goal is an integration time of about 1 millisecond, leading to 1,000 data pages per second. If there are 1 million pixels per data page, then the readout rate is 1 gigapixel per second. In the HRAM system this data rate can be supported continuously, independent of readout order. The latency is dominated by the integration time and is typically about 1 millisecond. This can be reduced by using a pulsed laser or a CMOS detector array or both.

Applications

The HRAM system is best matched to applications with high capacity and fast readout rate demands, yet with relatively infrequent changes to the stored data. Video-on-demand and Web servers fit well here: movie and Web content change infrequently, yet multiple users are continuously accessing enormous amounts of content in a fairly random order. (Playing one movie for a single user is sequential— playing 10 movies at once is not.)

An alternative method for reaching 1 Tbit in storage capacity is a jukebox-type apparatus. Blocks of material, each containing 1 Gbit or more, are brought into position in front of the reference beam optics for readout. The access time to a hologram is either 1 millisecond if the hologram is in the current material block, or several seconds if it is in a separate block. This can be reduced somewhat by having several readout stations. A principal advantage of increasing the capacity in this way is in the cost per megabyte of storage. For a one-block HRAM system, the cost is dominated by the components: camera, SLM, laser, beam steerers, and optics. The advantages provided by the lack of moving parts are probably enough to support this cost per megabyte only for military applications. For the commercial market, however, the cost per megabyte drops rapidly as more blocks are used, until the cost of the material becomes dominant.

Associative Retrieval

An HRAM system also lets designers use a unique feature of holographic storage: associative retrieval. To search a conventional storage device for all data records sharing a particular feature, we would retrieve each record into RAM, search it using software, and continue until all records were recalled and checked. With holographic storage, this process can be performed at the memory itself.

Instead of reconstructing signal data pages with a reference beam, the data pattern of interest is put on the SLM and illuminates the storage location with the signal beam. All the reference beams used to store holograms in that stack are reconstructed. The brightness of each beam, however, is proportional to the correlation between the original stored data pattern and the data pattern from the interrogating signal beam. Once the reference beams are focused onto a “correlation plane” detector array, the reference beam angles corresponding to the closest matches can be identified. The reference beam can then be used to reconstruct the desired data page onto the output camera, completing the search-and-retrieval process in perhaps 5 milliseconds.

Compact Modular Holographic Memory

One drawback in the HRAM system is that the number of rapidly accessible locations (and thus the immediately accessible capacity) is limited by the beam-steering optics. Rather than bring the beams to the storage material, another approach is to bring the pixel arrays for data input and output to the storage material. In fact, by applying a unique feature of the stored holograms, the same pixel array can be used for both input and output. Upon readout, instead of bringing back the same reference beam used during recording, its “phase conjugate” is directed to the storage location. (The phase conjugate of an optical beam passes backward along the beam’s path, like a movie played in reverse.) This new readout beam reconstructs the phase conjugate of the signal beam, which returns along the original signal path back to the SLM. Because of this, a phase conjugate signal beam allows the use of a cheap imaging lens, or even no lens at all. If each pixel of the SLM is not only a light modulator but also a detector, then the entire storage device can be fabricated from identical compact modules with no moving parts. One such module is shown in Figure 4. With several modules, the memory resembles a board of DRAM, with holography increasing the amount of data stored per RAM chip.

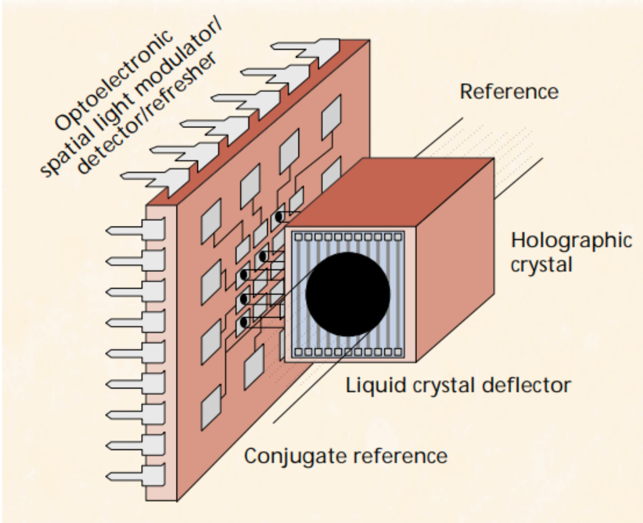


Figure 4. A compact modular holographic system can be formed by placing the holographic material close to a pixel array that performs both input and output.

Caltech researchers recently demonstrated a singlemodule compact holographic system with 480 modulator/detector pixels and 25 stored holograms. A small amount of logic at each “smart” pixel let the system counteract the erasure in a photorefractive crystal by periodically detecting and refreshing the holograms. This can remove the need to fix the stored holograms and makes it feasible to individually erase holograms from a stack.

In a modular system, the cost per megabyte is dominated by the smart pixel array and the two compact angle steerers (one for the writing beam, one for the readout). Associative retrieval is possible, but adds another detector array per module. Increasing the number of pixels while keeping the cost of the detector array and angle tuners low is the key to practical implementation of this architecture.

4.5.4 Holographic 3D disks

A third approach to spatial multiplexing leaves the optics and components stationary and moves the storage material. The simplest method to do this employs a disk configuration. The disk is constructed with a thick layer (approximately 1 mm) of the holographic material. Multiple holograms can be stored at each location on the disk surface. These locations are arranged along radial tracks, with the motion of the head selecting a track, and the disk rotation providing access along each track. As the medium thickness increases, the number of holograms that can be stored increases (and therefore the surface density goes up). However, the surface area that is illuminated also increases (and therefore the surface density goes down). Typically, it would be desirable to fabricate disks of 1-mm thickness, yielding a density of approximately 100 bits per squared micron. Even though the data is still stored in 3D, in the disk configuration the surface density, not the volume density, is what matters for most practical purposes.

Angle multiplexing can be used to multiplex holograms on the holographic disk in a manner similar to the HRAM architecture. However, the angle scanner would make the readout head too large and heavy for rapid access to holograms on different radial tracks. A single, simple reference beam that could attain the same density without a bulky beam deflector would be more convenient. This can be done by making the reference beam a spherical or converging beam. This is effectively equivalent to bringing in all the reference beam angles simultaneously instead of one at a time. In this case the reconstruction becomes very sensitive to the recorded hologram's position instead of the reference's angle of incidence. In fact, if the material is shifted by a few microns, the reconstruction disappears and a second hologram can be stored. The motion of the material relative to the illuminating spherical beam needed to reconstruct different shiftmultiplexed holograms is conveniently supplied by the disk's rotation. In addition, the simplicity of the reference beam makes the readout system look like a CD/DVD disk, albeit with parallel readout, as shown in Figure 5. The shift-multiplexing technique also depends on material thickness. Experimentally, it has been demonstrated that a thickness of 1 mm is sufficient to support 100 bits/μm^2, 20 times the areal density of a single-layer DVD.

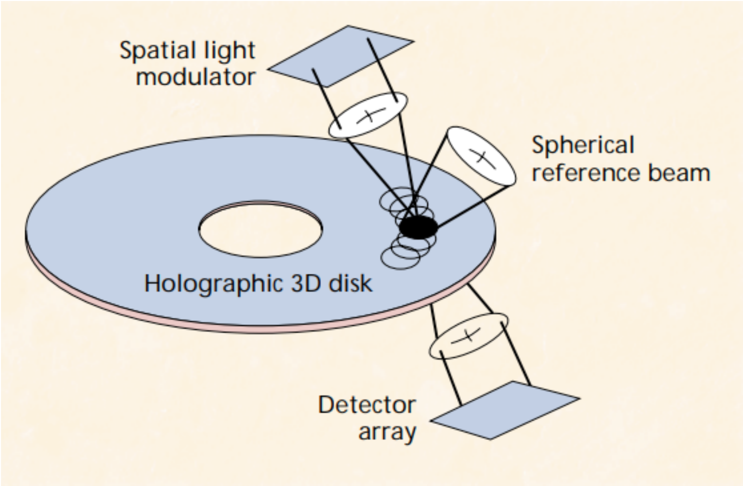


Figure 5. A shift-multiplexed disk resembles a compact disk, with an SLM and detector array for parallel input/output, and a simple reference beam. Track position and disk rotation enable access to densely overlapped holograms.

SUMMARY

- Holography is the science of making holograms which are usually intended for displaying three dimensional images. It is a physical structure that diffracts light into an image. A holographic image can be seen looking into an illuminated holographic print by shining a laser through a hologram and projecting an image on the screen.
- Optical holography needs a laser light to record the light field. In its early days, holography required high-power and expensive lasers, but currently, mass-produced low-cost laser diodes, such as those found on DVD recorders and used in other common applications, can be used to make holograms and have made holography much more accessible to low-budget researchers, artists and dedicated hobbyists.
- The hologram, that is, the medium which contains all the information, is nothing more than a high contrast, very fine grain, black and white photographic film. Silver halide emulsion much like the black and white film you can buy in your neighborhood drug store.
- A Denisyuk hologram is created with a 'single' beam' of laser light that is shone through the hologram surface and then bounces off object back into the hologram surface.
- Transmission holograms have a different a visual quality and the color is controlled by geometry rather than chemistry. When illuminated with a white (broad spectrum) light source the transmission hologram will diffract all the wavelengths of light into the image.
- Dot-matrix hologram printing is a technique of building up an image of diffractive 'pixels'. Each area is recorded with a particular geometry that diffracts light by a corresponding angle. The illuminating light is deflected into varying divergent color-spectrums.
- In laser holography, the hologram is recorded using a source of laser light, which is very pure in its color and orderly in its composition. Various setups may be used, and several types of holograms can be made, but all involve the interaction of light coming from different directions and producing a microscopic interference pattern which a plate, film, or other medium photographically records.

- A hologram can be made by shining part of the light beam directly into the recording medium, and the other part onto the object in such a way that some of the scattered light falls onto the recording medium. A more flexible arrangement for recording a hologram requires the laser beam to be aimed through a series of elements that change it in different ways.
- A hologram is created when a beam of laser light is split in two, with one of the resulting beams shone on an object (the object beam) then scattered onto a photographic plate, while the other beam is directed directly onto the plate (a reference beam).

REFERENCES

1. Beigzadeh, A.M. (2017). “Modelling of a holographic interferometry based calorimeter for radiation dosimetry”. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 864: 40–49.
2. Beigzadeh, A.M. (2018). “Double-exposure holographic interferometry for radiation dosimetry: A new developed model”. Radiation Measurements. 119: 132–139.
3. Blanche, P.-A.; Bablumian, A.; Voorakaranam, R.; Christenson, C.; Lin, W.; Gu, T.; Flores, D.; Wang, P.; et al. (2010). “Holographic three-dimensional telepresence using large-area photorefractive polymer”. Nature. 468 (7320): 80–83.
4. Gibney, E., “Physicist creates Star Wars-style 3D projections—just don’t call them holograms”, Nature, January 24, 2018.
5. Jones, A., et al., “Rendering for an Interactive 360 Light Field Display”, SIGGRAPH 2007 Papers Proceedings.
6. Kerrigan, S., “13 Hologram Projections Around the World That Are More Than Just a Pretty Display”, Interesting Engineering, May 17, 2018
7. MartíNez-Hurtado, J. L.; Davidson, C. A. B.; Blyth, J.; Lowe, C. R. (2010). “Holographic Detection of Hydrocarbon Gases and Other Volatile Organic Compounds”. Langmuir. 26 (19): 15694–15699.
8. Neverovich, M., “How Does Holographic Projection Work?”, Hypervsn, April 22, 2019.
9. Park, J., Lee, K., and Park, Y., “Ultrathin wide-angle large-area digital 3D holographic display using a non-periodic photon sieve”, Nature Communications 10, Article #1304, March 21, 2019.
10. Sainov, V.C., Stoykova, E.V, et al., “Trends in development of dynamic holographic displays,” SPIE Digital Library, June 9, 2006.
11. Stoykova, E., Kang, H., et al., “3D Capture and 3D Contents Generation for Holographic Imaging”, Chapter 8 of Holographic Materials and Optical Systems, Intech: 2017. DOI: 10.5772/65904

CHAPTER 5

LASER

INTRODUCTION

A laser is a device that emits light through a process of optical amplification based on the stimulated emission of electromagnetic radiation. The word «laser» is an acronym for «light amplification by stimulated emission of radiation". The first laser was built in 1960 by Theodore H. Maiman at Hughes Research Laboratories, based on theoretical work by Charles Hard Townes and Arthur Leonard Schawlow.

A laser differs from other sources of light in that it emits light which is *coherent*. Spatial coherence allows a laser to be focused to a tight spot, enabling applications such as laser cutting and lithography. Spatial coherence also allows a laser beam to stay narrow over great distances (collimation), enabling applications such as laser pointers and lidar. Lasers can also have high temporal coherence, which allows them to emit light with a very narrow spectrum. Alternatively, temporal coherence can be used to produce ultrashort pulses of light with a broad spectrum but durations as short as a femtosecond.

Lasers are used in optical disc drives, laser printers, barcode scanners, DNA sequencing instruments, fiber-optic, semiconducting chip manufacturing (photolithography), and free-space optical communication, laser surgery and skin treatments, cutting and welding materials, military and law enforcement devices for marking targets and measuring range and speed, and in laser lighting displays for entertainment. Semiconductor lasers in the blue to near-UV have also been used in place of light-emitting diodes (LED's) to excite fluorescence as a white light source. This permits a much smaller emitting area due to the much greater radiance of a laser and avoids the droop suffered by LED's; such devices are already used in some car headlamps.

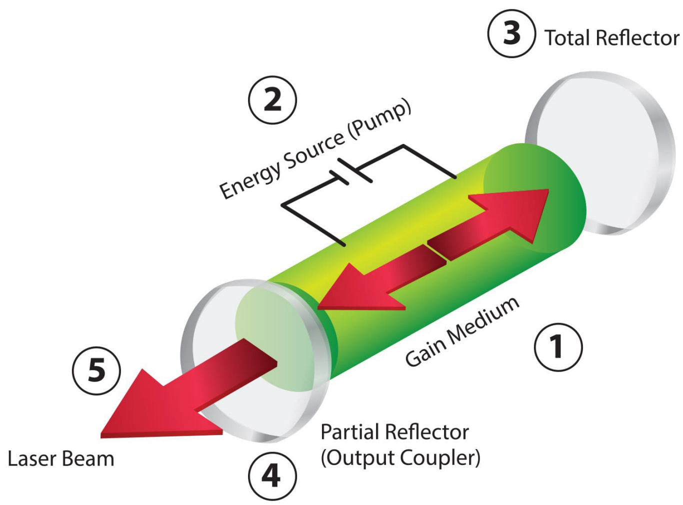
5.1 FUNDAMENTALS OF LASER

Lasers are distinguished from other light sources by their coherence. Spatial (or transverse) coherence is typically expressed through the output being a narrow beam, which is diffraction-limited. Laser beams can be focused to very tiny spots, achieving a very high irradiance, or they can have very low divergence in order to concentrate their power at a great distance. Temporal (or longitudinal) coherence implies a polarized wave at a single frequency, whose phase is correlated over a relatively great distance (the coherence length) along the beam. A beam produced by a thermal or other incoherent light source has an instantaneous amplitude and phase that vary randomly with respect to time and position, thus having a short coherence length.

Lasers are characterized according to their wavelength in a vacuum. Most "single wavelength" lasers actually produce radiation in several *modes* with slightly different wavelengths. Although temporal coherence implies some degree of monochromaticity, there are lasers that emit a broad spectrum of light or emit different wavelengths of light simultaneously. Some lasers are not single spatial mode and have light beams that diverge more than is required by the diffraction limit. All such devices are classified as "lasers" based on the method of producing light by stimulated emission. Lasers are employed where light of the required spatial or temporal coherence can not be produced using simpler technologies.

5.1.1 What is a Laser?

A laser is a device that emits a beam of coherent light through an optical amplification process. There are many types of lasers including gas lasers, fiber lasers, solid state lasers, dye lasers, diode lasers and excimer lasers. All of these laser types share a basic set of components.



5.1.2 Laser Components

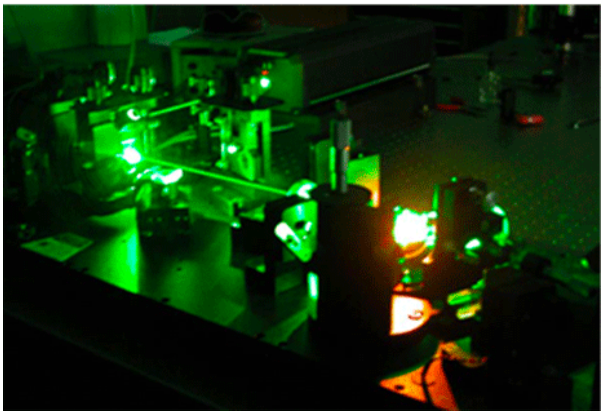
- Gain medium capable of sustaining stimulated emission
- Energy source to pump the gain medium
- Total reflector to reflect energy
- Partial reflector
- Laser beam output

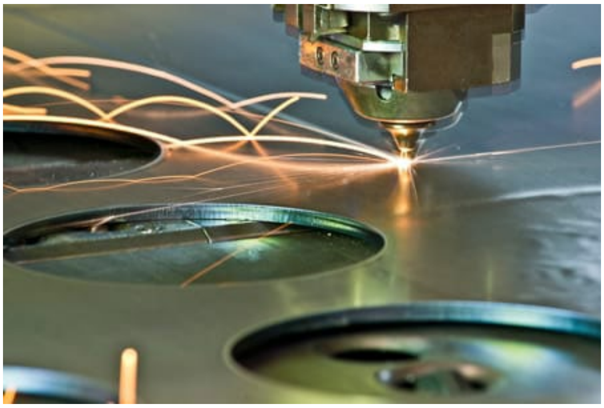
The gain medium and resonator determine the wavelength of the laser beam and the power of the laser.

5.1.3 How is Laser Technology Used?

Lasers are key components of many of the products that we use every day. Consumer products like Blu-Ray and DVD players rely on laser technology to read information from the disks. Bar code scanners rely on lasers for information processing. Lasers are also used in many surgical procedures such as LASIK eye surgery. In manufacturing, lasers are used for cutting, engraving, drilling and marking a broad range of materials.

- Laser Range Finding
- Information Processing (DVDs and Blu-Ray)
- Bar Code Readers
- Laser Surgery
- Holographic Imaging
- Laser Spectroscopy
- Laser Material Processing
    - Cutting
    - Engraving
    - Drilling
    - Marking
    - Surface Modification





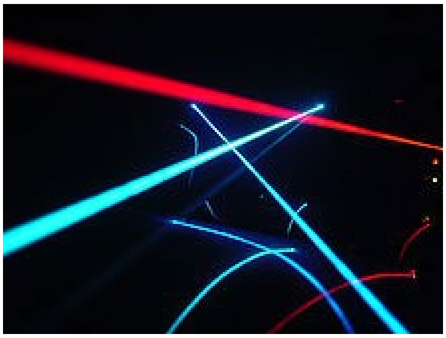
5.1.4 Terminology

The first device using amplification by stimulated emission operated at microwave frequencies, and was named “maser”, an acronym for “microwave amplification by stimulated emission of radiation”. When similar optical devices were developed they were first known as “optical masers”, until “microwave” was replaced by “light” in the acronym.

All such devices operating at frequencies higher than microwaves are called lasers (including *infrared laser*, *ultraviolet laser*, *X-ray laser* and *gamma-ray laser*). All devices operating at microwave or lower radio frequencies are called masers.

A laser that produces light by itself is technically an optical oscillator rather than an optical amplifier as suggested by the acronym. It has been humorously noted that the acronym LOSER, for "light oscillation by stimulated emission of radiation", would have been more correct. With the widespread use of the original acronym as a common noun, optical amplifiers have come to be referred to as "laser amplifiers".

The back-formed verb *to lase* is frequently used in the field, meaning «to give off coherent light,» especially in reference to the gain medium of a laser; when a laser is operating it is said to be «lasing». The words *laser* and *maser* are also used in cases where there is a coherent state unconnected with any manufactured device, as in *astrophysical maser* and *atom laser*.



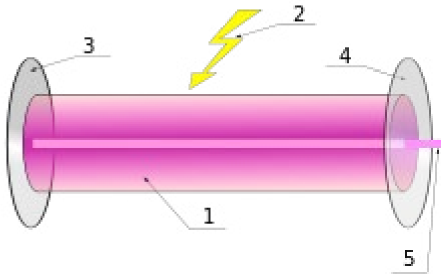
5.1.5 Design

A laser consists of a gain medium, a mechanism to energize it, and something to provide optical feedback. The gain medium is a material with properties that allow it to amplify light by way of stimulated emission. Light of a specific wavelength that passes through the gain medium is amplified (increases in power). Feedback enables stimulated emission to amplify predominantly the optical frequency at the peak of the gain-frequency curve. As stimulated emission grows, eventually one frequency dominates over all others, meaning that a coherent beam has been formed. The process of stimulated emission is analogous to that of an audio oscillator with positive feedback which can occur, for example, when the speaker in a public-address system is placed in proximity to the microphone. The screech one hears is audio oscillation at the peak of the gain-frequency curve for the amplifier.

For the gain medium to amplify light, it needs to be supplied with energy in a process called pumping. The energy is typically supplied as an electric current or as light at a different wavelength. Pump light may be provided by a flash lamp or by another laser.

The most common type of laser uses feedback from an optical cavity—a pair of mirrors on either end of the gain medium. Light bounces back and forth between the mirrors, passing through the gain medium and being amplified each time. Typically one of the two mirrors, the output coupler, is partially transparent. Some of the light escapes through this mirror. Depending on the design of the cavity (whether the mirrors are flat or curved), the light coming out of the laser may spread out or form a narrow beam. In analogy to electronic oscillators, this device is sometimes called a *laser oscillator*.

Most practical lasers contain additional elements that affect properties of the emitted light, such as the polarization, wavelength, and shape of the beam.



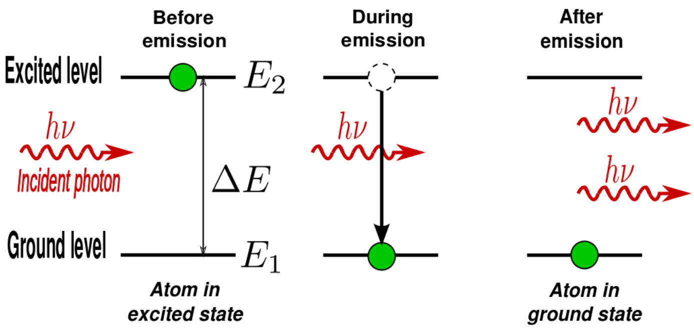
Components of a typical laser:

- Gain medium
- Laser pumping energy
- High reflector
- Output coupler
- Laser beam

5.1.6 Laser Physics

Electrons and how they interact with electromagnetic fields are important in our understanding of chemistry and physics.

In the classical view, the energy of an electron orbiting an atomic nucleus is larger for orbits further from the nucleus of an atom. However, quantum mechanical effects force electrons to take on discrete positions in orbitals. Thus, electrons are found in specific energy levels of an atom, two of which are shown below:



$$E_2 - E_1 = \Delta E = h\nu$$

An electron in an atom can absorb energy from light (photons) or heat (phonons) only if there is a transition between energy levels that matches the energy carried by the photon or phonon. For light, this means that any given transition will only absorb one particular wavelength of light. Photons with the correct wavelength can cause an electron to jump from the lower to the higher energy level. The photon is consumed in this process.

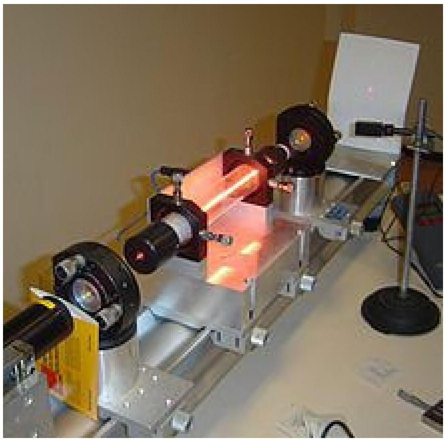
When an electron is excited from one state to that at a higher energy level with energy difference ΔE, it will not stay that way forever. Eventually, a photon will be spontaneously created from the vacuum having energy ΔE . Conserving energy, the electron transitions to a lower energy level which is not occupied, with transitions to different levels having different time constants. This process is called "spontaneous emission". Spontaneous emission is a quantum-mechanical effect and a direct physical manifestation of the Heisenberg uncertainty principle. The emitted photon has random direction, but its wavelength matches the absorption wavelength of the transition. This is the mechanism of fluorescence and thermal emission.

A photon with the correct wavelength to be absorbed by a transition can also cause an electron to drop from the higher to the lower level, emitting a new photon. The emitted photon exactly matches the original photon in wavelength, phase, and direction. This process is called stimulated emission.

Gain medium and cavity

The gain medium is put into an excited state by an external source of energy. In most lasers this medium consists of a population of atoms which have been excited into such a state by means of an outside light source, or an electrical field which supplies energy for atoms to absorb and be transformed into their excited states.

The gain medium of a laser is normally a material of controlled purity, size, concentration, and shape, which amplifies the beam by the process of stimulated emission described above. This material can be of any state: gas, liquid, solid, or plasma. The gain medium absorbs pump energy, which raises some electrons into higher-energy ("excited") quantum states. Particles can interact with light by either absorbing or emitting photons. Emission can be spontaneous or stimulated. In the latter case, the photon is emitted in the same direction as the light that is passing by. When the number of particles in one excited state exceeds the number of particles in some lower-energy state, population inversion is achieved. In this state, the rate of stimulated emission is larger than the rate of absorption of light in the medium, and therefore the light is amplified. A system with this property is called an optical amplifier. When an optical amplifier is placed inside a resonant optical cavity, one obtains a laser.



For lasing media with extremely high gain, so-called superluminescence, it is possible for light to be sufficiently amplified in a single pass through the gain medium without requiring a resonator. Although often referred to as a laser (see for example nitrogen laser), the light output from such a device lacks the spatial and temporal coherence achievable with lasers. Such a device cannot be described as an oscillator but rather is a high gain optical amplifier which amplifies its own spontaneous emission. The same mechanism describes so-called astrophysical masers/lasers.

The optical resonator is sometimes referred to as an "optical cavity", but this is a misnomer: lasers use open resonators as opposed to the literal cavity that would be employed at microwave frequencies in a maser. The resonator typically consists of two mirrors between which a coherent beam of light travels in both directions, reflecting back on itself so that an average photon will pass through the gain medium repeatedly before it is emitted from the output aperture or lost to diffraction or absorption. If the gain (amplification) in the medium is larger than the resonator losses, then the power of the recirculating light can rise exponentially. But each stimulated emission event returns an atom from its excited state to the ground state, reducing the gain of the medium. With increasing beam power the net gain (gain minus loss) reduces to unity and the gain medium is said to be saturated. In a continuous wave (CW) laser, the balance of pump power against gain saturation and cavity losses produces an equilibrium value of the laser power inside the cavity; this equilibrium determines the operating point of the laser. If the applied pump power is too small, the gain will never be sufficient to overcome the cavity losses, and laser light will not be produced. The minimum pump power needed to begin laser action is called the *lasing threshold*. The gain medium will amplify any photons passing through it, regardless of direction; but only the photons in a spatial mode supported by the resonator will pass more than once through the medium and receive substantial amplification.

The light emitted

In most lasers, lasing begins with spontaneous emission into the lasing mode. This initial light is then amplified by stimulated emission in the gain medium. Stimulated emission produces light that matches the input signal in direction, wavelength, and polarization, whereas the phase of emitted light is 90 degrees in lead of the stimulating light. This, combined with the filtering effect of the optical resonator gives laser light its characteristic coherence, and may give it uniform polarization and monochromaticity, depending on the resonator's design. The fundamental laser linewidth of light emitted from the lasing resonator can be orders of magnitude narrower than the linewidth of light emitted from the passive resonator. Some lasers use a separate injection seeder to start the process off with a beam that is already highly coherent. This can produce beams with a narrower spectrum than would otherwise be possible.

In 1963, Roy J. Glauber showed that coherent states are formed from combinations of photon number states, for which he was awarded the Nobel Prize in physics. A coherent beam of light is formed by single-frequency quantum photon states distributed according to a Poisson distribution. As a result, the arrival rate of photons in a laser beam is described by Poisson statistics.

Many lasers produce a beam that can be approximated as a Gaussian beam; such beams have the minimum divergence possible for a given beam diameter. Some lasers, particularly high-power ones, produce multimode beams, with the transverse modes often approximated using Hermite–Gaussian or Laguerre-Gaussian functions. Some high power lasers use a flat-topped profile known as a "tophat beam". Unstable laser resonators (not used in most lasers) produce fractal-shaped beams. Specialized optical systems can produce more complex beam geometries, such as Bessel beams and optical vortexes.

Near the "waist" (or focal region) of a laser beam, it is highly *collimated*: the wavefronts are planar, normal to the direction of propagation, with no beam divergence at that point. However, due to diffraction, that can only remain true well within the Rayleigh range. The beam of a single transverse mode (gaussian beam) laser eventually diverges at an angle which varies inversely with the beam diameter, as required by diffraction theory. Thus, the "pencil beam" directly generated by a common helium–neon laser would spread out to a size of perhaps 500 kilometers when shone on the Moon (from the distance of the earth). On the other hand, the light from a semiconductor laser typically exits the tiny crystal with a large divergence: up to 50°. However even such a divergent beam can be transformed into a similarly collimated beam by means of a lens system, as is always included, for instance, in a laser pointer whose light originates from a laser diode. That is possible due to the light being of a single spatial mode. This unique property of laser light, spatial coherence, cannot be replicated using standard light sources (except by discarding most of the light) as can be appreciated by comparing the beam from a flashlight (torch) or spotlight to that of almost any laser.

A laser beam profiler is used to measure the intensity profile, width, and divergence of laser beams.

Diffuse reflection of a laser beam from a matte surface produces a speckle pattern with interesting properties.

Quantum vs. classical emission processes

The mechanism of producing radiation in a laser relies on stimulated emission, where energy is extracted from a transition in an atom or molecule. This is a quantum phenomenon that was predicted by Albert Einstein, who derived the relationship between the A coefficient describing spontaneous emission and the B coefficient which applies to absorption and stimulated emission. However, in the case of the free electron laser, atomic energy levels are not involved; it appears that the operation of this rather exotic device can be explained without reference to quantum mechanics.

5.1.7 Continuous and Pulsed Modes of Operation

A laser can be classified as operating in either continuous or pulsed mode, depending on whether the power output is essentially continuous over time or whether its output takes the form of pulses of light on one or another time scale. Of course even a laser whose output is normally continuous can be intentionally turned on and off at some rate in order to create pulses of light. When the modulation rate is on time scales much slower than the cavity lifetime and the time period over which energy can be stored in the lasing medium or pumping mechanism, then it is still classified as a “modulated” or “pulsed” continuous wave laser. Most laser diodes used in communication systems fall in that category.

Continuous wave operation

Some applications of lasers depend on a beam whose output power is constant over time. Such a laser is known as *continuous wave* (*CW*). Many types of lasers can be made to operate in continuous wave mode to satisfy such an application. Many of these lasers actually lase in several longitudinal modes at the same time, and beats between the slightly different optical frequencies of those oscillations will, in fact, produce amplitude variations on time scales shorter than the round-trip time (the reciprocal of the frequency spacing between modes), typically a few nanoseconds or less. In most cases, these lasers are still termed “continuous wave” as their output power is steady when averaged over any longer time periods, with the very high-frequency power variations having little or no impact in the intended application. (However, the term is not applied to mode-locked lasers, where the *intention* is to create very short pulses at the rate of the round-trip time.)

For continuous wave operation, it is required for the population inversion of the gain medium to be continually replenished by a steady pump source. In some lasing media, this is impossible. In some other lasers, it would require pumping the laser at a very high continuous power level which would be impractical or destroy the laser by producing excessive heat. Such lasers cannot be run in CW mode.

Pulsed operation

Pulsed operation of lasers refers to any laser not classified as continuous wave, so that the optical power appears in pulses of some duration at some repetition rate. This encompasses a wide range of technologies addressing a number of different motivations. Some lasers are pulsed simply because they cannot be run in continuous mode.

In other cases, the application requires the production of pulses having as large an energy as possible. Since the pulse energy is equal to the average power divided by the repetition rate, this goal can sometimes be satisfied by lowering the rate of pulses so that more energy can be built up in between pulses. In laser ablation, for example, a small volume of material at the surface of a work piece can be evaporated if it is heated in a very short time, while supplying the energy gradually would allow for the heat to be absorbed into the bulk of the piece, never attaining a sufficiently high temperature at a particular point.

Other applications rely on the peak pulse power (rather than the energy in the pulse), especially in order to obtain nonlinear optical effects. For a given pulse energy, this requires creating pulses of the shortest possible duration utilizing techniques such as Q-switching.

The optical bandwidth of a pulse cannot be narrower than the reciprocal of the pulse width. In the case of extremely short pulses, that implies lasing over a considerable bandwidth, quite contrary to the very narrow bandwidths typical of CW lasers. The lasing medium in some *dye lasers* and *vibronic solid-state lasers* produces optical gain over a wide bandwidth, making a laser possible which can thus generate pulses of light as short as a few femtoseconds (10^{-15} s).

Q-switching

In a Q-switched laser, the population inversion is allowed to build up by introducing loss inside the resonator which exceeds the gain of the medium; this can also be described as a reduction of the quality factor or ‘Q’ of the cavity. Then, after the pump energy stored in the laser medium has approached the maximum possible level, the introduced loss mechanism (often an electro- or acousto-optical element) is rapidly removed (or that occurs by itself in a passive device), allowing lasing to begin which rapidly obtains the stored energy in the gain medium. This results in a short pulse incorporating that energy, and thus a high peak power.

Mode-locking

A mode-locked laser is capable of emitting extremely short pulses on the order of tens of picoseconds down to less than 10 femtoseconds. These pulses will repeat at the round trip time, that is, the time that it takes light to complete one round trip between the mirrors comprising the resonator. Due to the Fourier limit (also known as energy-time uncertainty), a pulse of such short temporal length has a spectrum spread over a considerable bandwidth. Thus such a gain medium must have a gain bandwidth sufficiently broad to amplify those frequencies. An example of a suitable material is titanium-doped, artificially grown sapphire (Ti:sapphire) which has a very wide gain bandwidth and can thus produce pulses of only a few femtoseconds duration.

Such mode-locked lasers are a most versatile tool for researching processes occurring on extremely short time scales (known as femtosecond physics, femtosecond chemistry and ultrafast science), for maximizing the effect of nonlinearity in optical materials (e.g. in second-harmonic generation, parametric down-conversion, optical parametric oscillators and the like). Unlike the giant pulse of a Q-switched laser, consecutive pulses from a mode-locked laser are phase-coherent, that is, the pulses (and not just their envelopes) are identical and perfectly periodic. For this reason, and the extremely large peak powers attained by such short pulses, such lasers are invaluable in certain areas of research.

Pulsed pumping

Another method of achieving pulsed laser operation is to pump the laser material with a source that is itself pulsed, either through electronic charging in the case of flash lamps, or another laser which is already pulsed. Pulsed pumping was historically used with dye lasers where the inverted population lifetime of a dye molecule was so short that a high energy, fast pump was needed. The way to overcome this problem was to charge up large capacitors which are then switched to discharge through flashlamps, producing an intense flash. Pulsed pumping is also required for three-level lasers in which the lower energy level rapidly becomes highly populated preventing further lasing until those atoms relax to the ground state. These lasers, such as the excimer laser and the copper vapor laser, can never be operated in CW mode.

5.2 PRINCIPLES OF WORKING OF A LASER

In lasers, photons are interacted in three ways with the atoms:

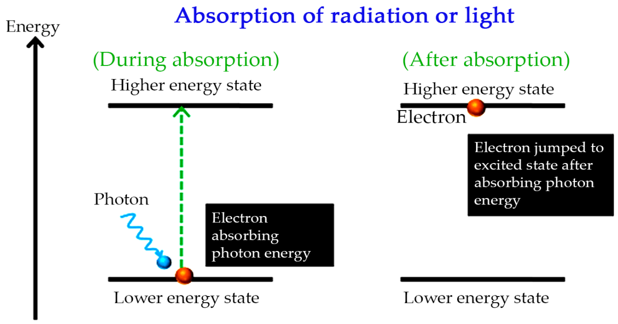
- Absorption of radiation
- Spontaneous emission
- Stimulated emission

5.2.1 Absorption of Radiation

Absorption of radiation is the process by which electrons in the ground state absorbs energy from photons to jump into the higher energy level.

The electrons orbiting very close to the nucleus are at the lower energy level or lower energy state whereas the electrons orbiting farther away from the nucleus are at the higher energy level. The electrons in the lower energy level need some extra energy to jump into the higher energy level. This extra energy is provided from various energy sources such as heat, electric field, or light.

Let us consider two energy levels (E_1 and E_2) of electrons. E_1 is the ground state or lower energy state of electrons and E_2 is the excited state or higher energy state of electrons. The electrons in the ground state are called lower energy electrons or ground state electrons whereas the electrons in the excited state are called higher energy electrons or excited electrons.



In general, the electrons in the lower energy state can't jump into the higher energy state. They need sufficient energy in order jump into the higher energy state.

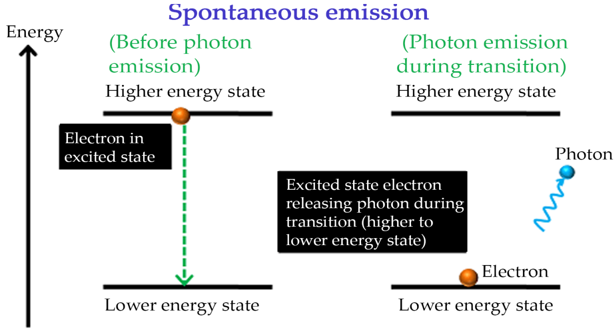
When photons or light energy equal to the energy difference of the two energy levels ($E_2 - E_1$) is incident on the atom, the ground state electrons gains sufficient energy and jumps from ground state (E_1) to the excited state (E_2).

The absorption of radiation or light occurs only if the energy of incident photon exactly matches the energy difference of the two energy levels ($E_2 - E_1$).

5.2.2 Spontaneous Emission

Spontaneous emission is the process by which electrons in the excited state return to the ground state by emitting photons.

The electrons in the excited state can stay only for a short period. The time up to which an excited electron can stay at higher energy state (E_2) is known as the lifetime of excited electrons. The lifetime of electrons in excited state is 10^{-8} second.



Thus, after the short lifetime of the excited electrons, they return to the lower energy state or ground state by releasing energy in the form of photons.

In spontaneous emission, the electrons move naturally or spontaneously from one state (higher energy state) to another state (lower energy state) so the emission of photons also occurs naturally. Therefore, we have no control over when an excited electron is going to lose energy in the form of light.

The photons emitted in spontaneous emission process constitute ordinary incoherent light. Incoherent light is a beam of photons with frequent and random changes of phase between them. In other words, the photons emitted in the spontaneous emission process do not flow exactly in the same direction of incident photons.

5.2.3 Stimulated Emission

Stimulated emission is the process by which incident photon interacts with the excited electron and forces it to return to the ground state.

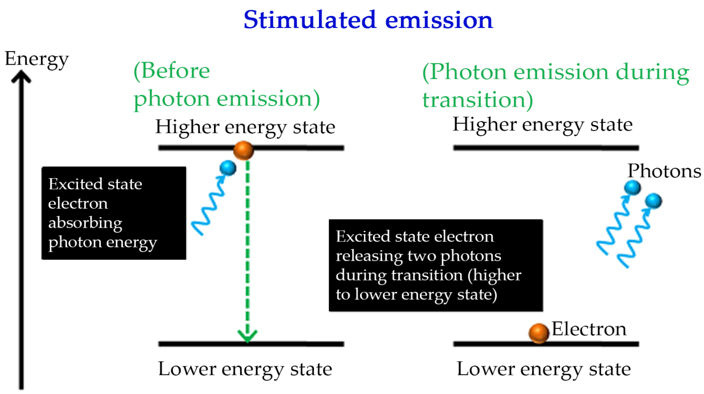
In stimulated emission, the light energy is supplied directly to the excited electron instead of supplying light energy to the ground state electrons.

Unlike the spontaneous emission, the stimulated emission is not a natural process it is an artificial process.

In spontaneous emission, the electrons in the excited state will remain there until its lifetime is over. After completing their lifetime, they return to the ground state by releasing energy in the form of light.

However, in stimulated emission, the electrons in the excited state need not wait for completion of their lifetime. An alternative technique is used to forcefully return the excited electron to ground state before completion of their lifetime. This technique is known as the stimulated emission.

When incident photon interacts with the excited electron, it forces the excited electron to return to the ground state. This excited electron release energy in the form of light while falling to the ground state.



In stimulated emission, two photons are emitted (one additional photon is emitted), one is due to the incident photon and another one is due to the energy release of excited electron. Thus, two photons are emitted.

The stimulated emission process is very fast compared to the spontaneous emission process.

All the emitted photons in stimulated emission have the same energy, same frequency and are in phase. Therefore, all photons in the stimulated emission travel in the same direction.

The number of photons emitted in the stimulated emission depends on the number of electrons in the higher energy level or excited state and the incident light intensity.

It can be written as:

Number of emitted photons α Number of electrons in the excited state + incident light intensity.

5.3 CHARACTERISTICS OF LASER

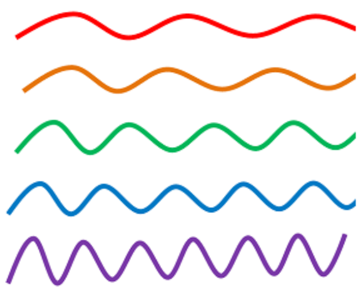
Laser light has four unique characteristics that differentiate it from ordinary light: these are

- Coherence
- Directionality
- Monochromatic
- High intensity

5.3.1 Coherence

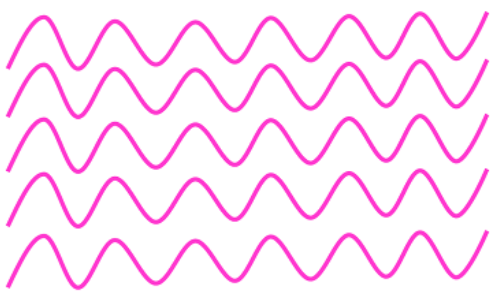
We know that visible light is emitted when excited electrons (electrons in higher energy level) jumped into the lower energy level (ground state). The process of electrons moving from higher energy level to lower energy level or lower energy level to higher energy level is called electron transition.

In ordinary light sources (lamp, sodium lamp and torch light), the electron transition occurs naturally. In other words, electron transition in ordinary light sources is random in time. The photons emitted from ordinary light sources have different energies, frequencies, wavelengths, or colors. Hence, the light waves of ordinary light sources have many wavelengths. Therefore, photons emitted by an ordinary light source are out of phase.



Incoherent light

In laser, the electron transition occurs artificially. In other words, in laser, electron transition occurs in specific time. All the photons emitted in laser have the same energy, frequency, or wavelength. Hence, the light waves of laser light have single wavelength or color. Therefore, the wavelengths of the laser light are in phase in space and time. In laser, a technique called stimulated emission is used to produce light.



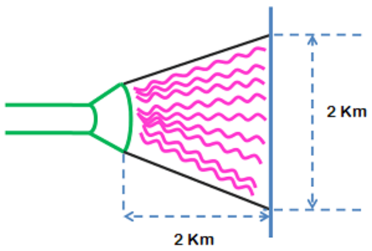
Coherent light waves

Thus, light generated by laser is highly coherent. Because of this coherence, a large amount of power can be concentrated in a narrow space.

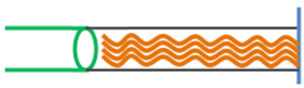
5.3.2 Directionality

In conventional light sources (lamp, sodium lamp and torchlight), photons will travel in random direction. Therefore, these light sources emit light in all directions.

On the other hand, in laser, all photons will travel in same direction. Therefore, laser emits light only in one direction. This is called directionality of laser light. The width of a laser beam is extremely narrow. Hence, a laser beam can travel to long distances without spreading.



Ordinary light



Laser light

If an ordinary light travels a distance of 2 km, it spreads to about 2 km in diameter. On the other hand, if a laser light travels a distance of 2 km, it spreads to a diameter less than 2 cm.

5.3.3 Monochromatic

Monochromatic light means a light containing a single color or wavelength. The photons emitted from ordinary light sources have different energies, frequencies, wavelengths, or colors. Hence, the light waves of ordinary light sources have many wavelengths or colors. Therefore, ordinary light is a mixture of waves having different frequencies or wavelengths.

On the other hand, in laser, all the emitted photons have the same energy, frequency, or wavelength. Hence, the light waves of laser have single wavelength or color. Therefore, laser light covers a very narrow range of frequencies or wavelengths.

5.3.4 High Intensity

You know that the intensity of a wave is the energy per unit time flowing through a unit normal area. In an ordinary light source, the light spreads out uniformly in all directions.

If you look at a 100 Watt lamp filament from a distance of 30 cm, the power entering your eye is less than 1/1000 of a watt.

In laser, the light spreads in small region of space and in a small wavelength range. Hence, laser light has greater intensity when compared to the ordinary light.

If you look directly along the beam from a laser (caution: don't do it), then all the power in the laser would enter your eye. Thus, even a 1 Watt laser would appear many thousand times more intense than 100 Watt ordinary lamp.

Thus, these four properties of laser beam enable us to cut a huge block of steel by melting. They are also used for recording and reproducing large information on a compact disc (CD).

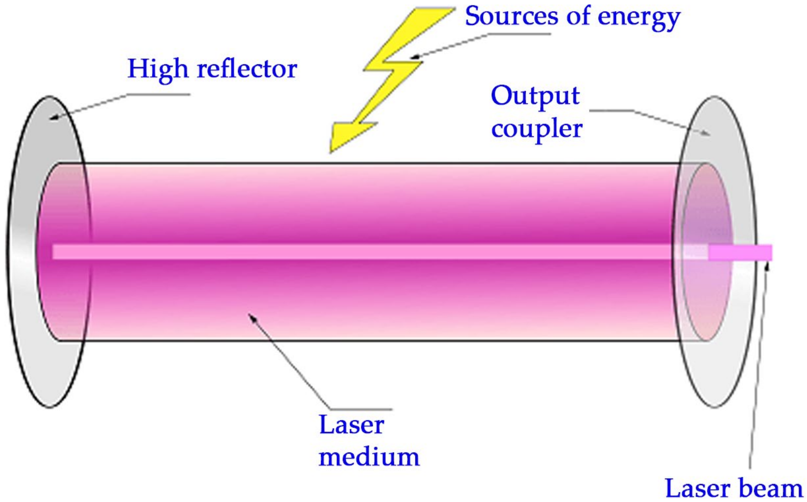
5.4 LASER CONSTRUCTION

A laser or laser system consists of three important components: a pump source, laser medium and optical resonator.

5.4.1 Pump Source

The pump source or energy source is the part of a laser system that provides energy to the laser medium. To get laser emission, first we need to produce population inversion. Population inversion is the process of achieving greater number of electrons in higher energy state as compared to the lower energy state.

The source of energy supplies sufficient amount of energy to the laser medium by which the electrons in the lower energy state are excited to the higher energy state. As a result, we get population inversion in the active medium or laser medium. Examples of energy sources include electric discharges, light from another laser, chemical reactions, and flash lamps. The type of energy source used is mostly depends on the laser medium. Excimer laser uses chemical reaction as energy source, a helium laser uses an electric discharge as energy source and Nd:YAG laser uses light focused from diode laser as energy source.



5.4.2 Laser Medium

The laser medium is a medium where spontaneous and stimulated emission of radiation takes place. Generally, the population of lower energy state is greater than the higher energy state. However, after achieving population inversion, the population of higher energy state becomes greater than the lower energy state.

After receiving sufficient energy from source, the electrons in the lower energy state or ground state are excited to the higher energy state (in the laser medium). The electrons in the excited state do not stay for long period because the lifetime of electrons in the excited state is very small. Hence, after a short period, the electrons in the excited state will fall back to the ground state by releasing energy in the form of light or photons. This is called spontaneous emission. In spontaneous emission, each electron emits a single photon while falling to the ground state.

When these emitted photons collide with the electrons in the excited state or meta stable state, it forces meta stable electrons to fall back to the ground state. As a result, electrons again release energy in the form of photons. This is called stimulated emission. In stimulated emission, each electron emits two photons while falling to the ground state.

When these emitted photons are again interacted with the meta stable state electrons then again two photons are emitted by each electron. Thus, millions of photons are generated by using only a small number of photons.

If we use electrical energy as energy source then a single photon or few photons (which are produced spontaneously) will produce large number of photons by stimulated emission process. Thus, light amplification is achieved in laser medium. Laser medium is also known as active medium or gain medium.

The laser medium will determine the characteristics of the laser light emitted. The laser medium can be solid, liquid, or gaseous.

Ruby laser is an example for solid-state laser. In this, a ruby crystal is used as an active medium. In this laser, xenon discharge tube which provides a flash light acts as pump source.

Helium – Neon laser is an example for gaseous laser. In this, neon is used as an active medium. In this laser, radio frequency (RF) generator acts as pump source.

5.4.3 Optical Resonator

The laser medium is surrounded by two parallel mirrors which provides feedback of the light. One mirror is fully reflective (100 % reflective) whereas another one is partially reflective (<100 % reflective). These two mirrors as a whole is called optical resonator. Optical resonator is also known as optical cavity or resonating cavity.

These two mirrors are given optical coatings which determine their reflective properties. Optical coating is a thin layer of material deposited on materials such as mirror or lens. Each mirror is coated differently. Therefore, each mirror will reflect the light differently. One mirror will completely reflect the light whereas another one will partially reflect the light.

The completely reflective mirror is called high reflector whereas the partially reflective mirror is called output coupler. The output coupler will allows some of the light to leave the optical cavity to produce the laser's output beam.

When energy is supplied to the laser medium, the lower energy state electrons in the laser medium will moves to excited state. After a short period, the electrons in the excited state will fall back to the ground state by releasing energy in the form of photons or light. This process of emission of photons is called spontaneous emission. Thus, light is produced in an active medium by a process called spontaneous emission.

The light generated within the laser medium will bounce back and forth between the two mirrors. This stimulates other electrons to release light while falling to the ground state. Likewise, a large number of electrons are stimulated to emit light. Thus, optical gain is achieved.

This amplified light escapes through the partially reflecting mirror. The process of stimulating electrons of other atoms to produce light in the laser medium is called stimulated emission.

The light in the laser medium is reflected many hundreds of times between the mirrors before it escape through the partially reflecting mirror. The light escaped from the partially reflecting mirror is produced by the stimulated emission process. Hence, this light will travel to large distances without spreading in the space.

5.5 TYPES OF LASERS

LASER stands for Light Amplification by Stimulated Emission of Radiation. A laser is a device which produces highly directional light. It emits light through a process called stimulated emission of radiation which increases the intensity of light.

A laser is different from conventional light sources in four ways: coherence, directionality, monochromacity, and high intensity.

The light waves of ordinary light sources have many wavelengths. Hence, the photons emitted by ordinary light sources are out of phase. Thus, ordinary light is incoherent.

On the other hand, the light waves of laser light have only one wavelength. Hence, all the photons emitted by laser light are in phase. Thus, laser light is coherent.

The light waves from laser contain only one wavelength or color so it is known as monochromatic light.

The laser beam is very narrow and can be concentrated on a very small area. This makes laser light highly directional.

The laser light spreads in a small region of space. Hence, all the energy is concentrated on a narrow region.Therefore, laser light has greater intensity than the ordinary light.

Lasers are classified into 4 types based on the type of laser medium used:

- Solid-state laser
- Gas laser
- Liquid laser
- Semiconductor laser

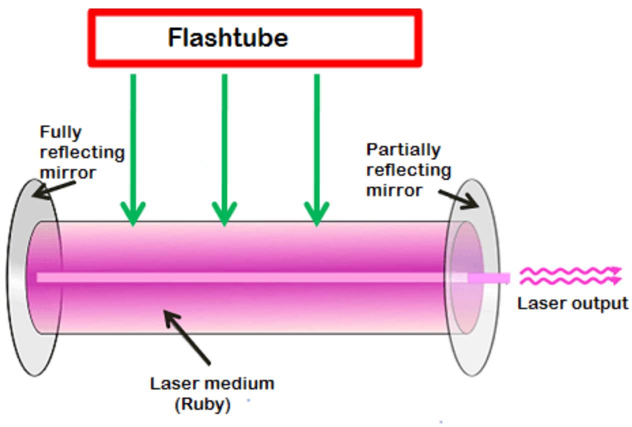
5.5.1 Solid-state Laser

A solid-state laser is a laser that uses solid as a laser medium. In these lasers, glass or crystalline materials are used.

Ions are introduced as impurities into host material which can be a glass or crystalline. The process of adding impurities to the substance is called doping. Rare earth elements such as cerium (Ce), erbium (Eu), terbium (Tb) etc are most commonly used as dopants.

Materials such as sapphire (Al_2O_3), neodymium-doped yttrium aluminum garnet (Nd:YAG), Neodymium-doped glass (Nd:glass) and ytterbium-doped glass are used as host materials for laser medium. Out of these, neodymium-doped yttrium aluminum garnet (Nd:YAG) is most commonly used.

The first solid-state laser was a ruby laser. It is still used in some applications. In this laser, a ruby crystal is used as a laser medium.

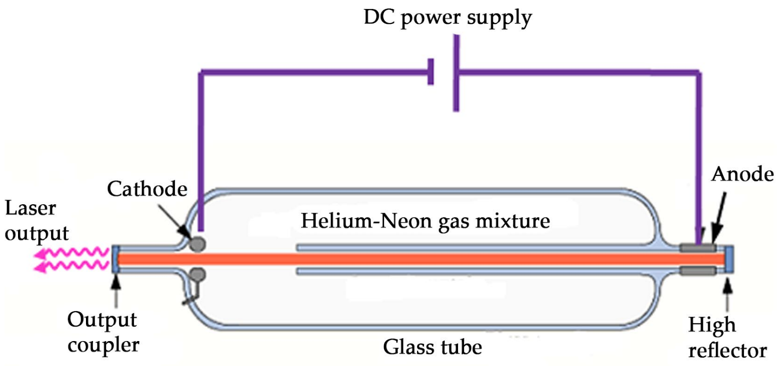


In solid-state lasers, light energy is used as pumping source. Light sources such as flashtube, flash lamps, arc lamps, or laser diodes are used to achieve pumping.

Semiconductor lasers do not belong to this category because these lasers are usually electrically pumped and involve different physical processes.

5.5.2 Gas Laser

A gas laser is a laser in which an electric current is discharged through a gas inside the laser medium to produce laser light. In gas lasers, the laser medium is in the gaseous state.



Gas lasers are used in applications that require laser light with very high beam quality and long coherence lengths.

In gas laser, the laser medium or gain medium is made up of the mixture of gases. This mixture is packed up into a glass tube. The glass tube filled with the mixture of gases acts as an active medium or laser medium.

A gas laser is the first laser that works on the principle of converting electrical energy into light energy. It produces a laser light beam in the infrared region of the spectrum at 1.15 μm.

Gas lasers are of different types: they are, Helium (He) – Neon (Ne) lasers, argon ion lasers, carbon dioxide lasers (CO_2 lasers), carbon monoxide lasers (CO lasers), excimer lasers, nitrogen lasers, hydrogen lasers, etc. The type of gas used to construct the laser medium can determine the lasers wavelength or efficiency.

5.5.3 Liquid Laser

A liquid laser is a laser that uses the liquid as laser medium. In liquid lasers, light supplies energy to the laser medium.

A dye laser is an example of the liquid laser. A dye laser is a laser that uses an organic dye (liquid solution) as the laser medium.

A dye laser is made up of an organic dye mixed with a solvent. These lasers generate laser light from the excited energy states of organic dyes dissolved in liquid solvents. It produces laser light beam in the near ultraviolet (UV) to the near infrared (IR) region of the spectrum.

5.5.4 Semiconductor Laser

Semiconductor lasers play an important role in our everyday life. These lasers are very cheap, compact size and consume low power. Semiconductor lasers are also known as laser diodes.

Semiconductor lasers are different from solid-state lasers. In solid-state lasers, light energy is used as the pump source whereas, in semiconductor lasers, electrical energy is used as the pump source.

In semiconductor lasers, a p-n junction of a semiconductor diode forms the active medium or laser medium. The optical gain is produced within the semiconductor material.

5.5.5 Ruby Laser

Ruby laser definition

A ruby laser is a solid-state laser that uses the synthetic ruby crystal as its laser medium. Ruby laser is the first successful laser developed by Maiman in 1960.

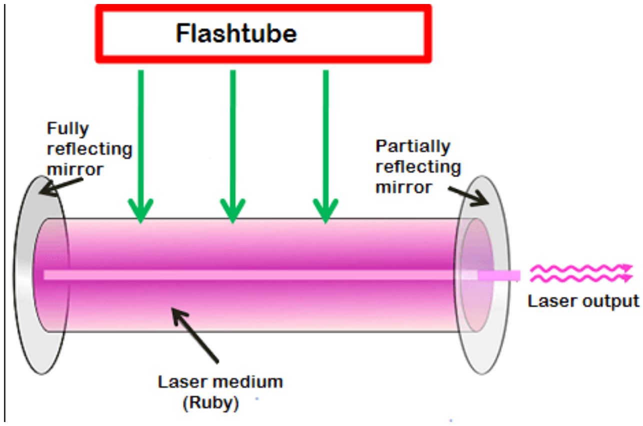
Ruby laser is one of the few solid-state lasers that produce visible light. It emits deep red light of wavelength 694.3 nm.

Construction of ruby laser

A ruby laser consists of three important elements: laser medium, the pump source, and the optical resonator.

Laser medium or gain medium in ruby laser

In a ruby laser, a single crystal of ruby ($Al_2O_3 : Cr^{3+}$) in the form of cylinder acts as a laser medium or active medium. The laser medium (ruby) in the ruby laser is made of the host of sapphire (Al_2O_3) which is doped with small amounts of chromium ions (Cr^{3+}). The ruby has good thermal properties.



Pump source or energy source in ruby laser

The pump source is the element of a ruby laser system that provides energy to the laser medium. In a ruby laser, population inversion is required to achieve laser emission. Population inversion is the process of achieving the greater population of higher energy state than the lower energy state. In order to achieve population inversion, we need to supply energy to the laser medium (ruby).

In a ruby laser, we use flashtube as the energy source or pump source. The flashtube supplies energy to the laser medium (ruby). When lower energy state electrons in the laser medium gain sufficient energy from the flashtube, they jump into the higher energy state or excited state.

Optical resonator

The ends of the cylindrical ruby rod are flat and parallel. The cylindrical ruby rod is placed between two mirrors. The optical coating is applied to both the mirrors. The process of depositing thin layers of metals on glass substrates to make mirror surfaces is called silvering. Each mirror is coated or silvered differently.

At one end of the rod, the mirror is fully silvered whereas, at another end, the mirror is partially silvered.

The fully silvered mirror will completely reflect the light whereas the partially silvered mirror will reflect most part of the light but allows a small portion of light through it to produce output laser light.

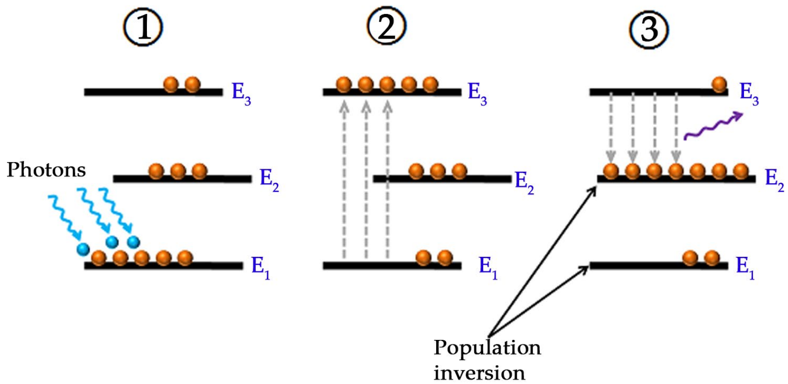
Working of ruby laser

The ruby laser is a three level solid-state laser. In a ruby laser, optical pumping technique is used to supply energy to the laser medium. Optical pumping is a technique in which light is used as energy source to raise electrons from lower energy level to the higher energy level.

Consider a ruby laser medium consisting of three energy levels E_1, E_2, E_3 with N number of electrons.

We assume that the energy levels will be $E_1 < E_2 < E_3$. The energy level E_1 is known as ground state or lower energy state, the energy level E_2 is known as metastable state, and the energy level E_3 is known as pump state.

Let us assume that initially most of the electrons are in the lower energy state (E_1) and only a tiny number of electrons are in the excited states (E_2 and E_3)



When light energy is supplied to the laser medium (ruby), the electrons in the lower energy state or ground state (E_1) gains enough energy and jumps into the pump state (E_3).

The lifetime of pump state E_3 is very small (10^{-8} sec) so the electrons in the pump state do not stay for long period. After a short period, they fall into the metastable state E_2 by releasing radiationless energy. The lifetime of metastable state E_2 is 10^{-3} sec which is much greater than the lifetime of pump state E_3. Therefore, the electrons reach E_2 much faster than they leave E_2. This results in an increase in the number of electrons in the metastable state E_2 and hence population inversion is achieved.

After some period, the electrons in the metastable state E_2 falls into the lower energy state E_1 by releasing energy in the form of photons. This is called spontaneous emission of radiation.

When the emitted photon interacts with the electron in the metastable state, it forcefully makes that electron fall into the ground state E_1. As a result, two photons are emitted. This is called stimulated emission of radiation.

When these emitted photons again interacted with the metastable state electrons, then 4 photons are produced. Because of this continuous interaction with the electrons, millions of photons are produced.

In an active medium (ruby), a process called spontaneous emission produces light. The light produced within the laser medium will bounce back and forth between the two mirrors. This stimulates other electrons to fall into the ground state by releasing light energy. This is called stimulated emission. Likewise, millions of electrons are stimulated to emit light. Thus, the light gain is achieved.

The amplified light escapes through the partially reflecting mirror to produce laser light.

5.5.6 Nd:YAG Laser

Nd:YAG laser definition

Neodymium-doped Yttrium Aluminum Garnet (Nd: YAG) laser is a solid state laser in which Nd: YAG is used as a laser medium.

These lasers have many different applications in the medical and scientific field for processes such as Lasik surgery and laser spectroscopy.

Nd: YAG laser is a four-level laser system, which means that the four energy levels are involved in laser action. These lasers operate in both pulsed and continuous mode.

Nd: YAG laser generates laser light commonly in the near-infrared region of the spectrum at 1064 nanometers (nm). It also emits laser light at several different wavelengths including 1440 nm, 1320 nm, 1120 nm, and 940 nm.

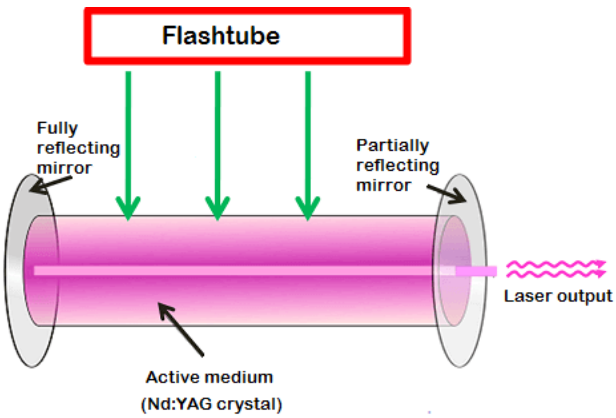
Nd: YAG laser construction

Nd:YAG laser consists of three important elements: an energy source, active medium, and optical resonator.

Energy source

The energy source or pump source supplies energy to the active medium to achieve population inversion. In Nd: YAG laser, light energy sources such as flashtube or laser diodes are used as energy source to supply energy to the active medium.

In the past, flashtubes are mostly used as pump source because of its low cost. However, nowadays, laser diodes are preferred over flashtubes because of its high efficiency and low cost.



Active medium

The active medium or laser medium of the Nd:YAG laser is made up of a synthetic crystalline material (Yttrium Aluminum Garnet (YAG)) doped with a chemical element (neodymium (Nd)). The lower energy state electrons of the neodymium ions are excited to the higher energy state to provide lasing action in the active medium.

Optical resonator

The Nd:YAG crystal is placed between two mirrors. These two mirrors are optically coated or silvered.

Each mirror is silvered or coated differently. One mirror is fully silvered whereas, another mirror is partially silvered. The mirror, which is fully silvered, will completely reflect the light and is known as fully reflecting mirror.

On the other hand, the mirror which is partially silvered will reflect most part of the light but allows a small portion of light through it to produce the laser beam. This mirror is known as a partially reflecting mirror.

Working of Nd:YAG laser

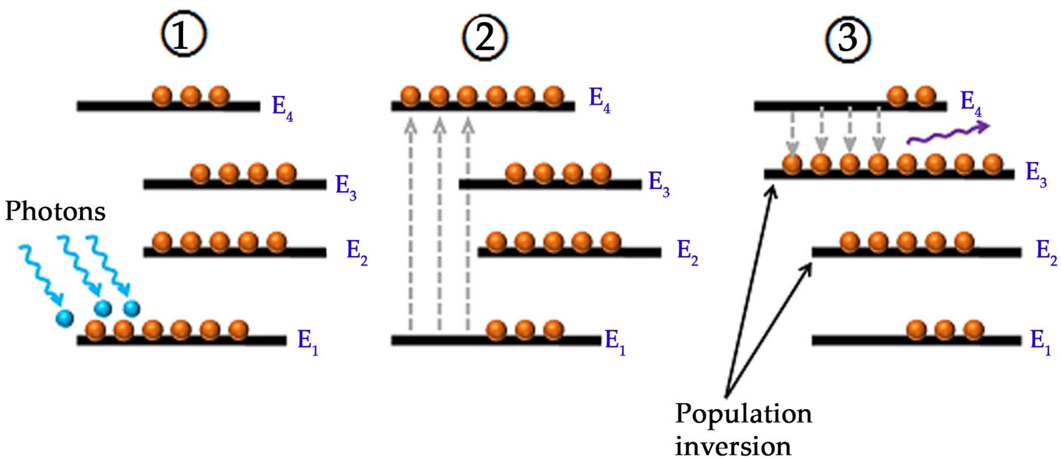
Nd: YAG laser is a four-level laser system, which means that the four energy levels are involved in laser action. The light energy sources such as flashtubes or laser diodes are used to supply energy to the active medium.

In Nd:YAG laser, the lower energy state electrons in the neodymium ions are excited to the higher energy state to achieve population inversion.

Consider a Nd:YAG crystal active medium consisting of four energy levels E_1, E_2, E_3, and E_4 with N number of electrons. The number of electrons in the energy states E1, E_2, E_3, and E_4 will be N_1, N_2, N_3, and N_4.

Let us assume that the energy levels will be $E_1 < E_2 < E_3 < E_4$. The energy level E_1 is known as ground state, E_2 is the next higher energy state or excited state, E_3 is the metastable state or excited state and E_4 is the pump state or excited state. Let us assume that initially, the population will be $N_1 > N_2 > N_3 > N_4$.

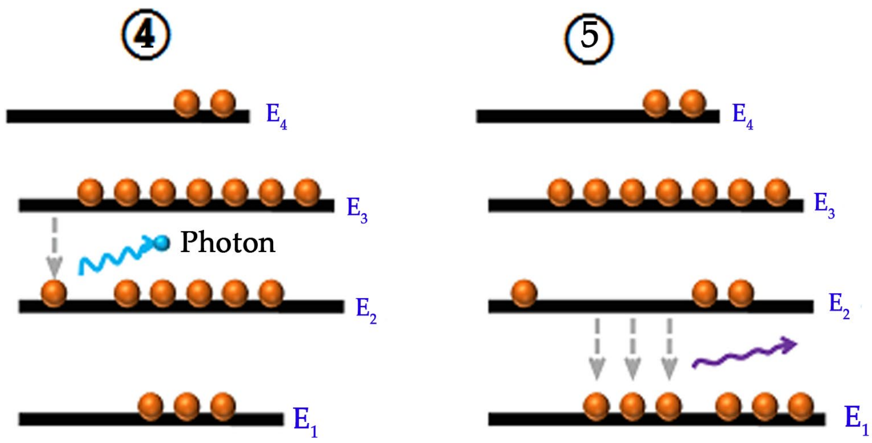
When flashtube or laser diode supplies light energy to the active medium (Nd:YAG crystal), the lower energy state (E_1) electrons in the neodymium ions gains enough energy and moves to the pump state or higher energy state E_4.



The lifetime of pump state or higher energy state E_4 is very small (230 microseconds (Âμs)) so the electrons in the energy state E_4 do not stay for long period. After a short period, the electrons will fall into the next lower energy state or metastable state E_3 by releasing non-radiation energy (releasing energy without emitting photons).

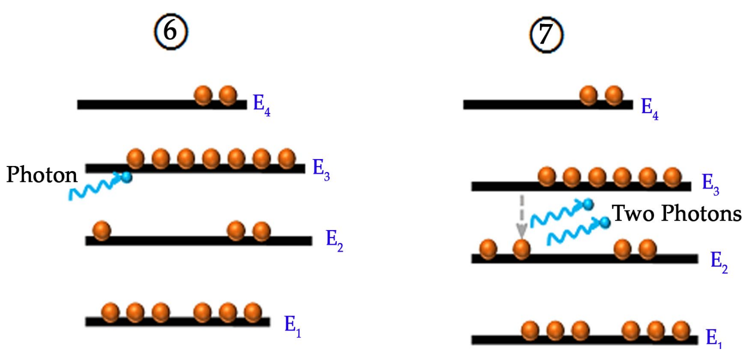
The lifetime of metastable state E_3 is high as compared to the lifetime of pump state E_4. Therefore, the electrons reach E_3 much faster than they leave E_3. This results in an increase in the number of electrons in the metastable E_3 and hence population inversion is achieved.

After some period, the electrons in the metastable state E_3 will fall into the next lower energy state E_2 by releasing photons or light. The emission of photons in this manner is called spontaneous emission.



The lifetime of energy state E_2 is very small just like the energy state E_4. Therefore, after a short period, the electrons in the energy state E_2 will fall back to the ground state E_1 by releasing radiationless energy.

When photon emitted due to spontaneous emission is interacted with the other metastable state electron, it stimulates that electron and makes it fall into the lower energy state by releasing the photon. As a result, two photons are released. The emission of photons in this manner is called stimulated emission of radiation.



When these two photons again interacted with the metastable state electrons, four photons are released. Likewise, millions of photons are emitted. Thus, optical gain is achieved.

Spontaneous emission is a natural process but stimulated emission is not a natural process. To achieve stimulated emission, we need to supply external photons or light to the active medium.

The Nd:YAG active medium generates photons or light due to spontaneous emission. The light or photons generated in the active medium will bounce back and forth between the two mirrors. This stimulates other electrons to fall into the lower energy state by releasing photons or light. Likewise, millions of electrons are stimulated to emit photons.

The light generated within the active medium is reflected many times between the mirrors before it escapes through the partially reflecting mirror.

Advantages of Nd:YAG laser

- Low power consumption
- Nd:YAG laser offers high gain.
- Nd:YAG laser has good thermal properties.
- Nd:YAG laser has good mechanical properties.
- The efficiency of Nd:YAG laser is very high as compared to the ruby laser.

Applications of Nd:YAG laser

Military

Nd:YAG lasers are used in laser designators and laser rangefinders. A laser designator is a laser light source, which is used to target objects for attacking. A laser rangefinder is a rangefinder, which uses a laser light to determine the distance to an object.

Medicine

Nd: YAG lasers are used to correct posterior capsular opacification (a condition that may occur after a cataract surgery).

Nd:YAG lasers are used to remove skin cancers.

Manufacturing

Nd:YAG lasers are used for etching or marking a variety of plastics and metals.

Nd:YAG lasers are used for cutting and welding steel.

5.6 APPLICATIONS OF LASERS

Laser is an optical device that generates intense beam of coherent monochromatic light by stimulated emission of radiation.

Laser light is different from an ordinary light. It has various unique properties such as coherence, monochromacity, directionality, and high intensity. Because of these unique properties, lasers are used in various applications.

The most significant applications of lasers include:

- Lasers in medicine
- Lasers in communications
- Lasers in industries
- Lasers in science and technology
- Lasers in military

5.6.1 Lasers in Medicine

- Lasers are used for bloodless surgery.
- Lasers are used to destroy kidney stones.
- Lasers are used in cancer diagnosis and therapy.
- Lasers are used for eye lens curvature corrections.
- Lasers are used in fiber-optic endoscope to detect ulcers in the intestines.
- The liver and lung diseases could be treated by using lasers.
- Lasers are used to study the internal structure of microorganisms and cells.
- Lasers are used to produce chemical reactions.
- Lasers are used to create plasma.
- Lasers are used to remove tumors successfully.

- Lasers are used to remove the caries or decayed portion of the teeth.
- Lasers are used in cosmetic treatments such as acne treatment, cellulite and hair removal.

5.6.2 Lasers in Communications

- Laser light is used in optical fiber communications to send information over large distances with low loss.
- Laser light is used in underwater communication networks.
- Lasers are used in space communication, radars and satellites.

5.6.3 Lasers in Industries

- Lasers are used to cut glass and quartz.
- Lasers are used in electronic industries for trimming the components of Integrated Circuits (ICs).
- Lasers are used for heat treatment in the automotive industry.
- Laser light is used to collect the information about the prefixed prices of various products in shops and business establishments from the bar code printed on the product.
- Ultraviolet lasers are used in the semiconductor industries for photolithography. Photolithography is the method used for manufacturing printed circuit board (PCB) and microprocessor by using ultraviolet light.
- Lasers are used to drill aerosol nozzles and control orifices within the required precision.

5.6.4 Lasers in Science and Technology

- A laser helps in studying the Brownian motion of particles.
- With the help of a helium-neon laser, it was proved that the velocity of light is same in all directions.
- With the help of a laser, it is possible to count the number of atoms in a substance.
- Lasers are used in computers to retrieve stored information from a Compact Disc (CD).

- Lasers are used to store large amount of information or data in CD-ROM.
- Lasers are used to measure the pollutant gases and other contaminants of the atmosphere.
- Lasers helps in determining the rate of rotation of the earth accurately.
- Lasers are used in computer printers.
- Lasers are used for producing three-dimensional pictures in space without the use of lens.
- Lasers are used for detecting earthquakes and underwater nuclear blasts.
- A gallium arsenide diode laser can be used to setup an invisible fence to protect an area.

5.6.5 Lasers in Military

- Laser range finders are used to determine the distance to an object.
- The ring laser gyroscope is used for sensing and measuring very small angle of rotation of the moving objects.
- Lasers can be used as a secretive illuminators for reconnaissance during night with high precision.
- Lasers are used to dispose the energy of a warhead by damaging the missile.
- Laser light is used in LIDAR's to accurately measure the distance to an object.

SUMMARY

- A laser is a device that emits light through a process of optical amplification based on the stimulated emission of electromagnetic radiation.
- Lasers are distinguished from other light sources by their coherence. Spatial (or transverse) coherence is typically expressed through the output being a narrow beam, which is diffraction-limited.
- Lasers are key components of many of the products that we use every day. Consumer products like Blu-Ray and DVD players rely on laser technology to read information from the disks.
- A laser consists of a gain medium, a mechanism to energize it, and something to provide optical feedback.
- All the photons emitted in laser have the same energy, frequency, or wavelength.
- A laser or laser system consists of three important components: a pump source, laser medium and optical resonator.
- A solid-state laser is a laser that uses solid as a laser medium. In these lasers, glass or crystalline materials are used.
- A gas laser is a laser in which an electric current is discharged through a gas inside the laser medium to produce laser light.
- A liquid laser is a laser that uses the liquid as laser medium. In liquid lasers, light supplies energy to the laser medium.
- A ruby laser is a solid-state laser that uses the synthetic ruby crystal as its laser medium.

REFERENCES

1. “For The First Time, A Laser That Shines Pure White”. Popular Science. March 18, 2019.
2. “Guide to Lasers”. Hobarts. Retrieved April 24, 2017.
3. Bertolotti, Mario (2015). Masers and Lasers: An Historical Approach (2nd ed.). CRC Press. pp. 89–91. ISBN 978-1-4822-1780-3. Retrieved March 15, 2016.
4. Csele, Mark (2004). Fundamentals of Light Sources and Lasers. Wiley. ISBN 0-471-47660-9.
5. Nishizawa, Jun-ichi (December 2009). “Extension of frequencies from maser to laser”. Proc Jpn Acad Ser B Phys Biol Sci. 85 (10): 454–465. Bibcode:2009PJAB...85..454N.
6. Taylor, Nick (2000). Laser: The Inventor, The Nobel Laureate, and The Thirty-Year Patent War. Simon & Schuster. ISBN 978-0684835150.

CHAPTER 6

SPATIAL LIGHT MODULATORS

INTRODUCTION

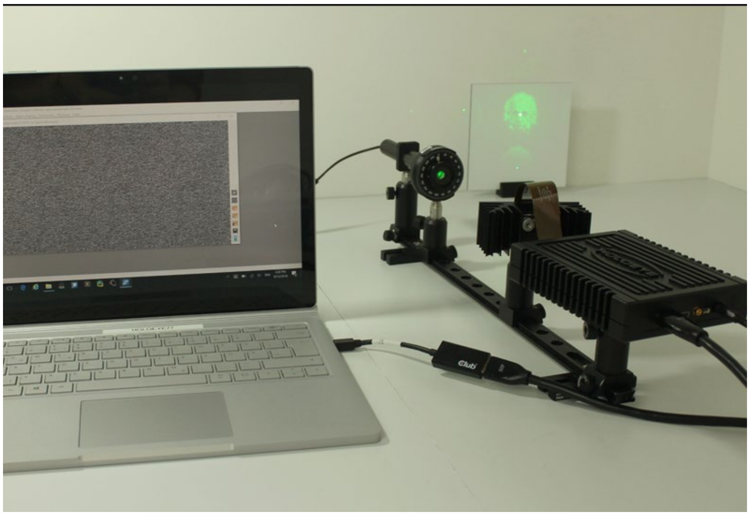
A spatial light modulator (SLM) is an object that imposes some form of spatially varying modulation on a beam of light. A simple example is an overhead projector transparency. Usually when the phrase SLM is used, it means that the transparency can be controlled by a computer. In the 1980s, large SLMs were placed on overhead projectors to project computer monitor contents to the screen. Since then more modern projectors have been developed where the SLM is built inside the projector. These are commonly used in meetings of all kinds for presentations. Usually, a SLM modulates the intensity of the light beam. However, it is also possible to produce devices that modulate the phase of the beam or both the intensity and the phase simultaneously. SLMs are used extensively in holographic data storage setups to encode information into a laser beam similarly to way a transparency does for an overhead projector. They can also be used as part of a holographic display technology. SLMs have been used as a component in optical computing. They also often find application in holographic optical tweezers. Liquid crystal SLMs can help solve problems related to laser micro particle manipulation. In this case spiral beam parameters can be changed dynamically.

6.1 OVERVIEW

Spatial light modulator (SLM) is a general term describing devices that are used to modulate amplitude, phase, or polarization of light waves in space and time. HOLOEYE´s Spatial Light Modulator systems are based on translucent (LCD) or reflective (LCOS) liquid crystal micro displays.

The use of LC materials in SLMs is based on their optical and electrical anisotropy. A certain gray level represents a defined average voltage across the LC cell. This voltage leads to a variable tilt of the LC molecules due to their electrical anisotropy. As LC molecules also show optical anisotropy this tilt changes the refractive index of the LC molecules (for suitable incident polarization, dependent on device version) which causes a modified optical path length within the LC cell. The addressed gray level is now converted into a phase level. HOLOEYEs SLMs are based on vertical aligned nematic (VAN), parallel aligned nematic (PAN) or twisted nematic (TN) micro display cells. In a twisted cell, the orientation of the molecules differs by typically 45°/90° between the top and the bottom of the LC cell and is arranged in a helix-like structure in between. In VAN / PAN cells the alignment layers are parallel to each other, so the LC molecules have the same orientation.

To use an SLM in amplitude modulation mode you need linear incident polarization. The transmitted or reflected light has to be guided through a 2nd polarizer (analyzer) that is crossed to the incident polarization. For phase modulation a setup without an analyzer is used. With devices based on twisted hematic LC or LCOS displays the twist always causes a polarization effect (amplitude modulation) and no phase only modulation is possible (phase mostly modulation). With VAN / PAN displays it is possible to modulate the phase without influence on the polarization / amplitude (phase only modulation) by using incident polarization along the LC director axis.



A spatial light modulator (SLM) is a special device that can manipulate light by modulating the amplitude, phase or polarization of the light waves in the two dimensions of space and time. This means that light is manipulated in order to obtain a desired output, and SLM is commonly used in overhead projectors such as those used in schools and office conference rooms.

Definitions of Specs

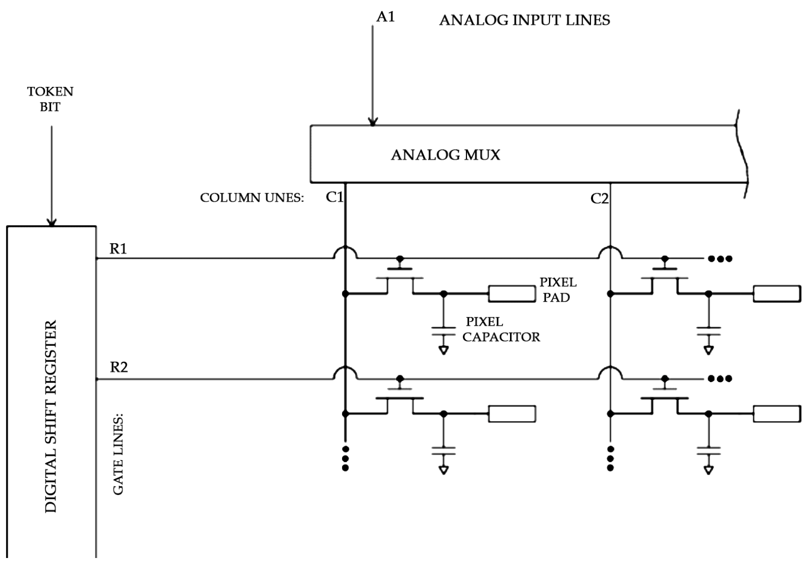
- **Resolution:** Number of pixels (width x height)
- **Pixel Pitch:** Size of a pixel including the interpixel gap
- **Fill Factor:** Surface area of the display which can actively use. There are gaps between the pixels at which the incident light is scattered.
- **Active Area:** Size of the actual addressable/usable display area.
- **Addressing:** Number of gray levels / phase levels that can be addressed. This can vary with addressing sequences.
- **Signal Formats:** Input signal format. Typically HDMI or DVI.
- **Input Frame Rate:** Addressing speed of the input signal (typically DVI / HDMI video frame rates of 60 Hz for monochrome applications).
- **Response Time:** The response time is defined as the switching time from 10% to 90% and from 90% to 10 % (rise and fall time). The actual response time of the liquid crystal is determined by the properties of the used liquid crystal material, the thickness of the LC layer, the used drive sequence / calibration (the actual

voltages applied to a pixel) and temperature. For phase SLMs the response time typically is below the Input frame rate.

- **Reflectivity:** Amount of light which is directly reflected (0-order of a non-addressed display). The reflectivity is not 100% as some of the light is diffracted into higher orders due to the grating like structure of the pixel matrix. Some part of the light is also scattered and absorbed at the interpixel gaps. In addition the reflectivity of the aluminum mirror is limited (dependent on wavelength).

Addressing the SLM

The optical function or information to be displayed on HOLOEYE Spatial Light Modulators can be taken directly from an optic design software or an image source and can be transferred by a computer interface. Using DVI/HDMI ports of standard PC graphics cards, the Spatial Light Modulator can be used just like an external plug & play monitor (e.g. in extended desktop mode for a second monitor) and no special software or drivers are necessary to operate the SLM. No matter which software is used to calculate optical functions, if you are able to address a bitmap image on a second monitor output of the graphics card you are able to operate the SLM. Also standard image viewer software can be used.



For an easy start and even more convenient operation HOLOEYE provides a software package with each SLM. It contains a Pattern Generator Software (for calculation of different optical functions), a Slideshow Player software (for easy addressing of recalculated functions or images on the

SLM) and an SLM Display SDK for different development environments. Also a Configuration Manager software is included for convenient configuration of the Spatial Light Modulator.

Possible Spatial Light Modulator Applications:

- Imaging & Projection
- Display Applications
- Holography (Display holography, holographic memory, holographic recording and security systems, including digital holography)
- Holographic Projection
- WSS – Wavelengths Selective Switching
- Beam Splitting
- Laser Beam Shaping
- Coherent Wave front Modulation
- Phase Shifting
- Optical Tweezers
- Laser Pulse Modulation

A spatial light modulator is an electronically programmable device that can modulate light output based on a specific fixed spatial pattern (pixel), essentially projecting light that is controlled in either amplitude only, phase only or both (phase-amplitude). This device makes use of liquid crystals to modulate the light, which is why overhead projectors are called LCD projectors.

There are many types of SLMs, and one common type is the electrically addressed SLM (EASLM), wherein the image is created and changed electronically just like in most electronic displays, and which usually receives input via conventional digital interfaces such as VGA or DVI. Another type is the optically addressed SLM (OASLM), which requires a separate light input encoded with an image that it can then project on its surface, again using liquid crystals. This means that an OASLM is a secondary display that takes input from an EASLM. In a process called image tiling, the images produced with an EASLM are then sequentially transferred to different parts of an OASLM before the whole image is displayed for the viewers. This can result in high-resolution images above 100 megapixels.

SLM Product Line-Up

Model No.	Features
SLM-300	High-Power LCOS Spatial Light Modulator
SLM-250	UV Hardened Spatial Light Modulator
SLM-200	Full HD Reflective Spatial Light Modulator

High-Power LCOS Spatial Light Modulator

SLM-300

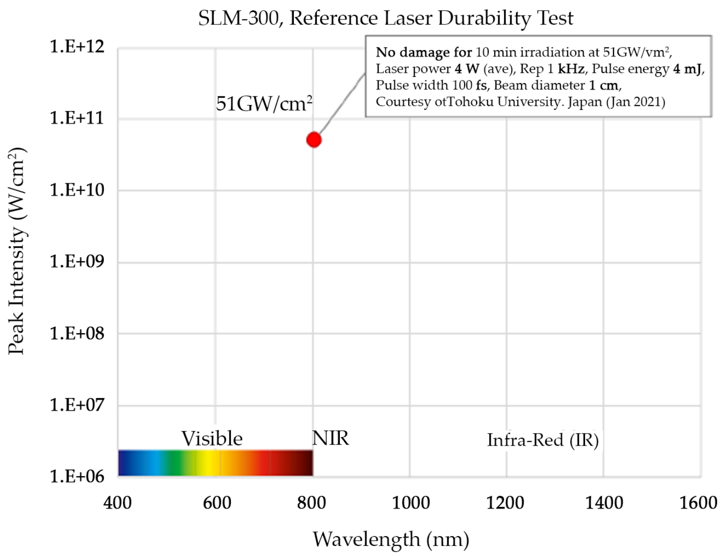
Our spatial light modulator (SLM) is based on reflective liquid crystal on silicon (LCOS) micro-display technology. The SLM-300 high-power model holds the world record for high-power durability up to CW 200 Watts. The SLM-300 has 100 times higher power durability when compared with Santec's conventional SLM product, the SLM-200.

【Features】

- High Power-durable LCOS based SLM
- High Phase Resolution 10-bit (1024) Gray Level
- Ultra Low Phase Noise ~0.003πrad.(Typ.)

【Application】

- Laser processing
- 3D printing
- IC trimming



6.2 PRINCIPLE AND APPLICATIONS

Liquid crystals are organic materials whose physico-chemical properties are intermediate between those of solids and liquids. The elongated LC molecules therefore have both a structural order and anisotropy specific to crystals such as optical, dielectric or even elastic anisotropy. The optical anisotropy is generally higher than in crystals: an optical birefringence ranging between 0.1 and 0.2 for example is common in LCs. The existence of a liquid order also guarantees the fluid nature of the different LC mesophases and, to some extent, the tunability of their properties. These characteristics, combined with a wide spectral transparency, have made LCs materials widely used in optics. In the nematic mesophase, molecules do not have a positional order but are oriented in a preferential direction, defined by a vector n $\vec{n}$, the so-called director axis. They feature properties of an anisotropic uniaxial medium with an optical axis oriented along n $\vec{n}$. This direction can be experimentally specified by defining specific boundary conditions at the surface of the sample. Typical anchoring conditions can be planar (PAN), homeotropic (VAN) or twisted (TN). Furthermore, the orientation of the director axis can be controlled by applying external electric and / or magnetic fields. As a matter of fact, the application of an electric field results in the creation of elastic forces leading to the reorientation of the molecules that tend to line up in the direction where the strain energy is minimal. For LCs with a positive dielectric anisotropy, the minimum energy is reached when the molecules are aligned in the direction of the electric field. The strength of the electric field modulates the average molecular orientation. Therefore, the optical refractive index of the medium is electrically-controlled and the phase of a propagating light is modified accordingly.

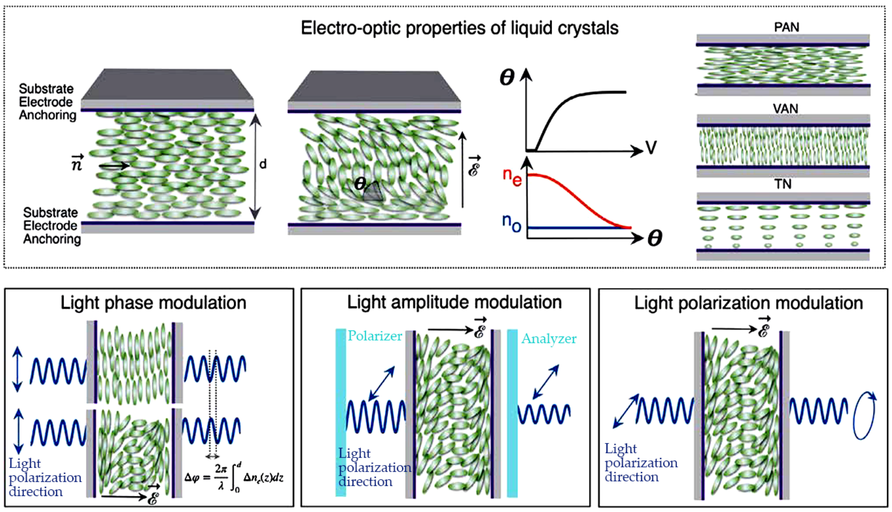


Figure 1: Electro-optical properties of nematic liquid crystal layers enable to locally change the phase of the propagating readout light.

The application of the electric field induces an average molecular rotation, which in return changes the refractive index, according to the input light polarization. Typical planar anchoring conditions can be vertical (VAN) or horizontal (PAN) or both, e.g. twisted (TN). Such a simple device allows for the modulation of the phase, amplitude or polarization of light according to the design details and the presence or absence of additional polarizing elements.

Spatial control of the applied electric field, on one or two dimensions, offers the ability to spatially modulate the phase of an incident optical wave. The latter is referred in the following as the “readout beam”, while the recording signal contains the information to be “printed” on the phase of the readout light. The readout beam has to be polarized. In addition, its polarization is a mean to control the parameter modulated by the LC component, whether it is the phase, amplitude or polarization. Indeed, projection of an initially linearly polarized light at 45° with respect to the LC extraordinary axis provides a phase-mismatch between the two crossed-polarized components. This variable phase shift allows the modulation of the light amplitude if the component is placed between polarizer and analyzer, and the modification of the linear polarization state to elliptical otherwise. Subsequently, phase-only, amplitude-only, polarization, or the combination of phase-amplitude modulation can be readily realized with a LC-SLM, as illustrated in Figure 1.

6.2.1 Applications

For decades, the large market of image projection and displays has fed the development of LC-SLMs that address otherwise uncountable applications in a wide field of scientific investigations. Nowadays, SLMs are used in fields as varied as imaging, digital holography, optical switching, microstructure fabrication, optical vortex generation. In the context of adaptive optics, SLMs are employed to correct the wave front of lasers and optimize the point spread function for biomedical applications and microscopy. In addition, SLMs enable ultra-short optical pulse shaping through a process known as Fourier domain pulse shaping. Recently, such devices have also been used in the field of telecommunications in order to achieve modal multiplexing in multimode optical fibers.

6.2.2 SLM Major Families and Performances

Most of commercial SLMs are electrically-addressed, for instance through standard digital video interface with each grey level being related to a given voltage. They can operate either in reflection or in transmission.

Transmissive SLMs

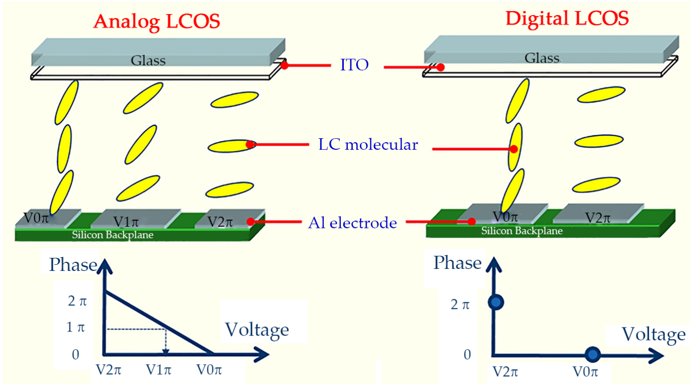
Transmissive SLMs include a nematic LC layer confined between two transparent conductive windows. At least one electrode is segmented, in order to provide individual electrical control over a certain amount of pixels. A particular type of transmissive SLM is the well-known LCD (Liquid Crystal Display), for amplitude light modulation. LCD relies on TN-type LC-SLM, placed between parallel or crossed polarizers.

Reflective SLMs

Reflective SLMs are particularly sought as they enable to fold the associated optical system, while light propagates twice in the modulating layer, which, in turn, increases the dynamic range. In this family, the most popular technology is LCoS. Liquid Crystal on Silicium, mainly used for phase-only or amplitudeonly light modulation. Most of current commercial electrically-addressed LCSLMs are based on this technology. LCoS are microdisplays, composed of a layer of nematic liquid crystals enclosed between a transparent electrode and a matrix of CMOS (complementary metal oxide semiconductor) integrated circuitry on a silicon backplane. The latter operate in reflection through a reflective treatment deposited on the CMOS matrix. Anchoring layers on one side of the electrode and on the reflective layer allow the molecules to be oriented in a direction parallel to the surface. An electric field maintained between the transparent electrode and the semiconductor controls the local average molecular orientation of the liquid crystal and modulates its refractive index.

Performances criteria

The different components of the LC-SLM multi-layer structure can be individually optimized according to the targeted application. Performances of LC-SLMs are then characterized as a priority by their active area, transmittance / reflectance, spectral acceptance, spatial resolution, response time and modulation dynamics.



Modulation range and response time

The modulation range is the maximum retardation that can be applied to a given wavelength while the dynamic response time is defined as the switching time from 10% to 90% and from 90% to 10% (rise and fall time). Theses two features are primarily determined by the LC layer material and thickness. Independently from the technology, the thickness of the LC layer is generally limited to 20 μm in most SLMs, and results from a balance between the desired modulation ranges (e.g. maximum phase modulation), maximum control voltage, molecular disorder issues, and dynamic response time. For radiation in the visible spectral range, the phase modulation evolves between 0 and 2π or 0 and 4π. The dynamic response time, meanwhile, ranges typically between 1–100 ms for 10–90% rise and fall times.

Active area

LC-SLMs for scientific applications present an active area usually around 1–2 cm_2, with some specific extension in the array configuration, up to 7 cm × 1 cm.

Spatial resolution

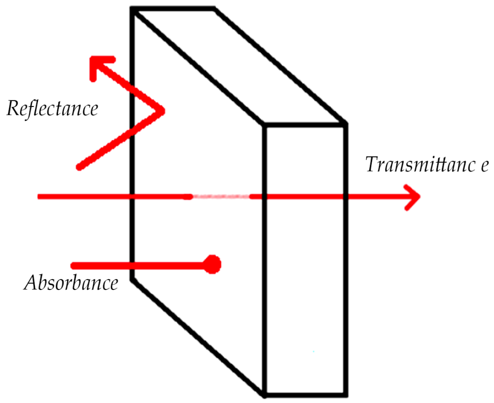
The spatial resolution is related to two parameters: the pixel density and the cross-talk between adjacent pixels. The pixel pitch depends on the category of SLMs. Higher pixel densities are achieved with LCoS, typically 1920 × 1080 pixels, but also up to 4160 × 2464. Transmissive SLMs are restricted to a larger pixel size (a few tens of μm for transmissive SLMs, as opposed to a few μm for LCoS). When specified beyond the simple number of pixels, the spatial resolution of an LC-SLM is around 40 lines per /mm, that is between 20–30 μm.

Filling factor

The electrically controlled LC-SLM makes it possible to control the properties of the readout light over a limited number of zones predefined by the manufacturer, this number being approximately equal to the number of electrode segments. In addition, at the junction between two adjacent segments, there are gaps where the SLM is inactive and / or has discontinuities in the modulated optical property. Inactive gaps are responsible for light scattering. This feature is translated as the so-called filling factor, usually slightly above 90% for commercial systems.

Reflectivity (transmittance)

The reflectivity (transmittance) is not 100% as some of the light may be diffracted into higher orders due to the grating like structure of the pixel matrix while some part of light is also scattered and absorbed at the interpixel gaps. In addition, the overall reflectivity (transmittance) is limited by losses at the multiple interfaces of the multi-layer structure, the electrode transparency, the reflectivity of the metallic or dielectric coating in LCoS. Typical values spread between 70% and 90%.



Spectral acceptance

Commercial SLMs make it possible to address different spectral ranges, with bandwidth around 200 nm, centered in the visible, nearinfrared or close to telecommunication bandwidths. A remaining drawback of the electrically-addressed LC-SLM technology is the presence of a top electrode. Most of the time, Indium Tin Oxyde (ITO, typically transparent over 0.3–1 μm, with partial transmission up to 1.5 μm) is employed, which tends to reduce the spectral acceptance, by comparison with the LC mixture itself. Moreover, as the phase modulation range scales with the optical frequency, extending the spectral range towards the infrared tends to require thicker LC layers, which, in turn, constraints the electrical addressing scheme. These two features limit the practical use of LC-SLMs in the visible and near infrared spectral range.

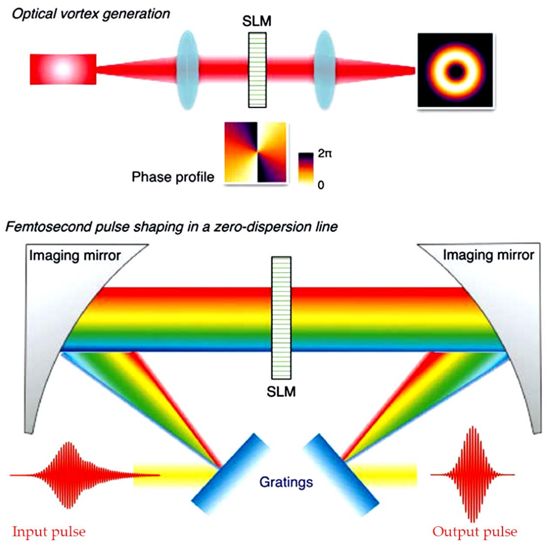


Figure 2: Illustration of some applications of SLMs. Top: a spiral phase pattern leads to optical vortex beams. Bottom: a SLM inserted in a zerodispersion line enables the temporal shaping of a femtosecond pulse.

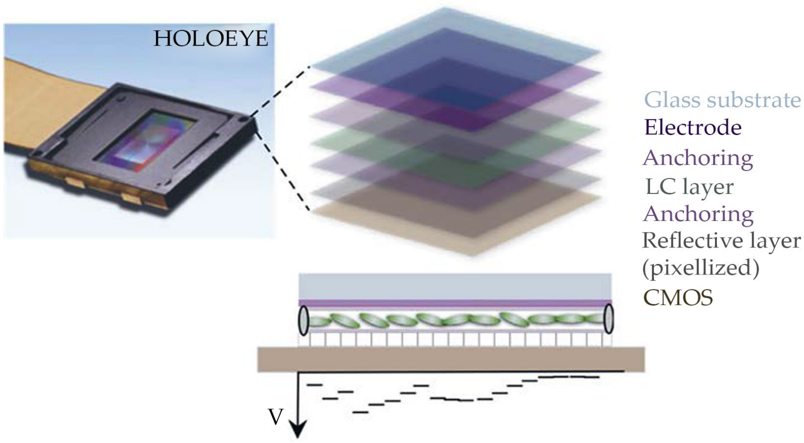


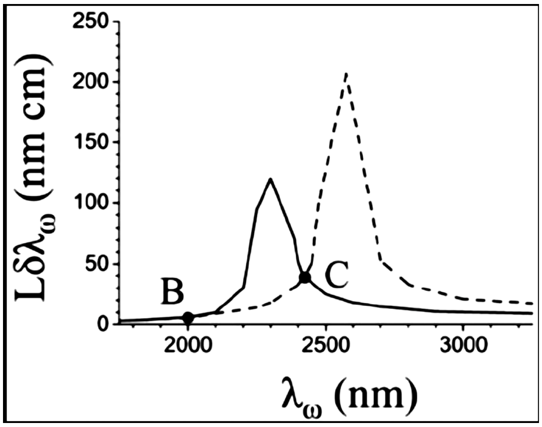
Figure 3: Structure and cross-section of an LCoS SLM and photograph of an industrial product from Holoeye.

Flicker

The flickering phase corresponds to the phase fluctuation due to electric polarization of the LC molecules and can be reduced to 0.01π by carefully designing control electronics.

Damage threshold

LC-SLM can tailor the properties of high power laser beams. Some damages might alter the SLM behavior, either due to laser ablation of one of the LC confining substrates, or to heating of the LC layer. According to the available data, the damage threshold is limited by electrodes and/or metallic coating in LCoS and is around 5W/ cm_2 for continuous light radiation, and decreases to 0.1 J/cm_2 for pulsed femtosecond lasers.



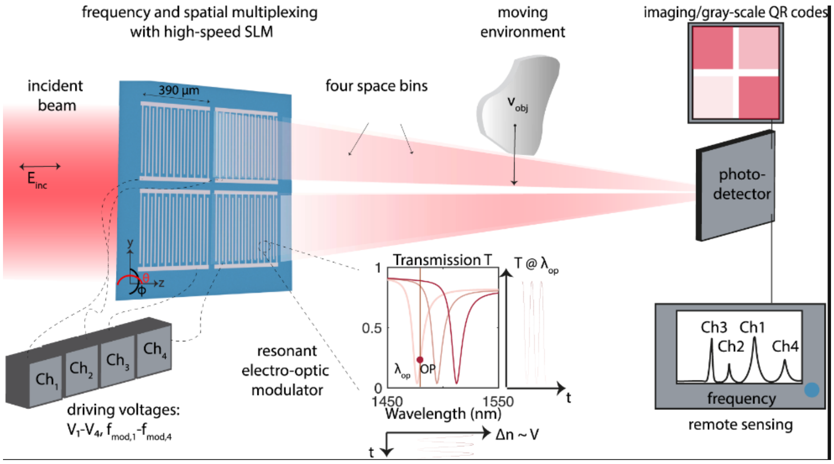
6.2.3 Limitations and Recent Advances

Although LC-SLMs are very performant and popular optical systems, some limitations can be deduced from the performances detailed above. Among them, the pixelisation and limited spectral acceptance in the mid-infrared spectral range are the most challenging issues. Nevertheless, research and development of innovative LC-SLMs is still very active. In particular, other solutions for controlling the birefringence are being investigated. Replacing the electrical addressing by an optical addressing solves the pixelisation issue and ensures arbitrary and continuous phase modulation. In so-called light valve modulators, or OASLM, a biased photo-conductive substrate replaces the segmented electrode and the voltage across the LC layer is locally controlled by an ancillary absorbed control beam, often referred as the “recording” beam, as opposed to the “readout” beam. This electrode is, however, mandatory, as an oscillating electric field has to be maintained across the LC layer to control the average orientation of the molecular director. Moreover, conventional optical valves require an isolation layer to prevent crosstalk between the recording and readout beams. Finally, novel technological developments in this field are steadily proposed. Thermal or thermooptical control of the LC layer or photo-polymerization of the anchoring layer might be promising methods to provide continuous phase modulation while eliminating the need for an electrode. Dielectric metasurfaces might also be part of the next-generation of SLMs.

6.3 APPLICATIONS OF SPATIAL LIGHT MODULATORS IN ATOM OPTICS

The application of optical fields in the area of atom optics is well established. Aside from the role of light beams in the standard cooling and trapping experiments that are routinely carried out in labs throughout the world they are also of interest for atom transport, atom guiding, evaporative cooling, dipole trapping and optical lattices.

There has also been significant progress in using magnetic fields to control and manipulate atoms, indeed magnetic trapping is a vital part of most realisations of Bose-Einstein Condensation. Most work on “atom-chip” devices is also reliant on the application of magnetic manipulation schemes. Such techniques have great power in that it is relatively straightforward to design fields that can move atoms in complicated, preordained geometries that are not conventionally possible with light fields. For instance, light fields cannot be bent in the manner of the magnetic work of Sauer *et al*. Magnetic fields have been used very successfully as atomic guides and beamsplitters and it is predicted that they will be useful in interferometers. We note that the magnetic potentials used in the interferometer proposed in need to be dynamically varied, something which is not as simple to do with light fields.



Despite the attractive possibilities of using magnetic potentials for atomic manipulation there do exist some drawbacks. Recent work by the Ketterle and Pritchard group and the Zimmermann group has shown that unexpected fragmentation of BECs occur when using magnetic guides due to geometric deformations of the current carrying wires and, possibly, other unexplained effects. Such deformations may pose limitations in atom-chip and magnetic waveguide physics. Other recent work has shown that there are limits for the coherent manipulation of atoms on atom chips due to the shot noise in current carrying wires.

Optical analogues to the magnetic atom chips have been proposed which make use of evanescent waves produced by integrated optical waveguides. Evanescent waves have already been shown to be a useful tool in trapped atoms above surfaces. Other possibilities include the use of micro fabricated optical elements to produce both individually addressable arrays of dipole traps and optical guides with applications in interferometry. Using such micro-optical devices it is possible to create tailored light potentials that will have a whole range of applications, but there are some drawbacks. Use of predetermined optical chips or light patterns are just that - predetermined. They require a certain amount of experimental hardware for each application that they are required to be used for, thus there are difficulties in reconfiguring set-ups. Furthermore such systems are not always suited for dynamic deformation of light potentials which may be useful for guiding, manipulating and interfering samples of cold atoms. In the following article we show how spatial light modulators (SLMs) can be used as powerful, highly reconfigurable, dynamically controllable holograms and their applications in the field of atom optics. Such devices can produce light patterns that require no micro fabrication and that can be projected into a vacuum system away from any surfaces. We demonstrate a number of light patterns that are difficult to generate using conventional micro-optical techniques but which can be made using etched holographic methods, although these lack the flexibility of the patterns generated using the SLM. These include a Mach-Zender interferometer, a Y-splitter, a bottle-beam and an array of dipole traps related to those generated by Dumke *et al* in. While such patterns can be generated using etched holograms, the SLM offers many advantages over such techniques.

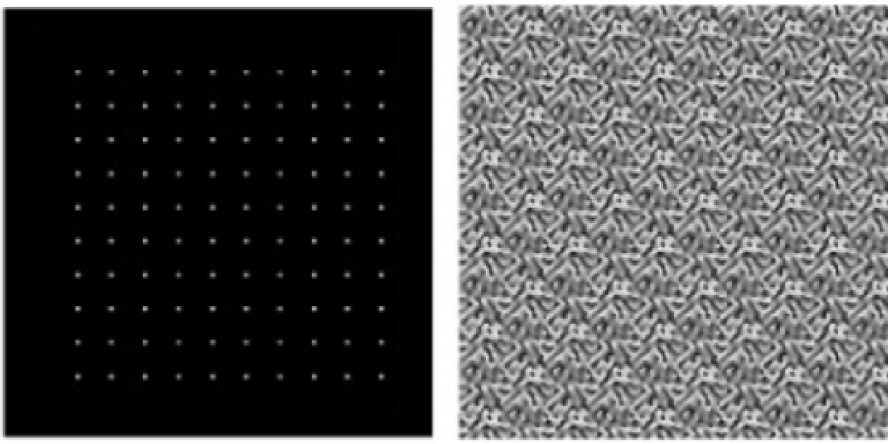
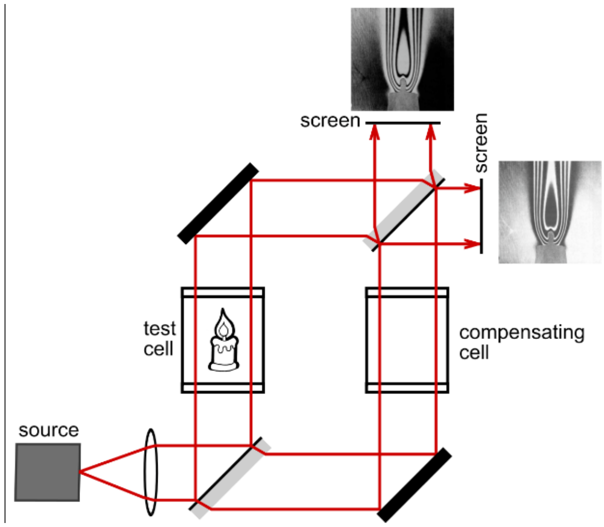


Figure 4: Holograms for a 10×10 square array of traps generated using the GS algorithm. (a) Input image (b) Generated phase hologram

6.3.1 Mach-Zender Interferometer

Atoms can be loaded into the light potential and then, due to intensity gradients within the light guides generated the atoms flow down the guides, subsequently being split and reformed. We are able to generate comparable potentials using the SLM and an example of a Mach-Zender interferometer-type structure is shown in cross-section in Fig. 5(a). We can image any pattern that we can generate into, say, a vacuum (as in) allowing flexible manipulation of the optical potential realized within the trap.

One of the attractions of magnetic guiding is that it is possible to design guides with curved geometry in which the atom cloud can be adiabatically transported. This is not so simple using current optical guiding techniques. The SLM generated patterns may allow this adiabatic transport within the optical regime.



6.3.2 Y-splitter

We are also able to make (as a simple extension of theMach-Zender pattern) a Y-splitter, which can be optimized to act as a 50:50 splitter (or indeed any other fraction). The splitter is shown in Fig. 5(b). A Y-splitter of this type is similar to that demonstrated in.

6.3.3 Blue-detuned Patterns

More difficult to create are hollow blue-detuned guides which would be useful as part of a blue-detuned interferometer, something not easily achievable using the micro-optics approach (although just as possible using etched holographic techniques).

A hollow Mach-Zender interferometer pattern is shown in Fig. 6(a). This is not explicitly generated as a pipe in three-dimensions, which is difficult using a phase only modulator, but it diffractively fills in, at least partially, on either side of the focus and hence forms an intensity null surrounded by high intensity barriers, necessary for a blue detuned guide. The filling process in this instance may not be enough however to constitute a stable atomic guiding potential, but such potentials should be possible with the inclusion of amplitude modulation effects, or with a more sophisticated holographic design technique.We estimate that the effective dipole potential of such a guide is 10mK (assuming a power of 100mW in the beam, a guide thickness of 0.01mm, a guide outer radius of

1.2mm and inner radius of 0.3mm. We also assume a guide detuning of 3GHz and that the atomic system in question is ^{85}Rb with a decay rate of 2π×6.1Mhz). We note that the generated image lies in a plane orthogonal to the optical axis of the SLM. To generate an image where the entrance to the guide lies parallel to the optical axis is, again, difficult using a phase-only SLM. However, this is a problem that can be overcome using a suitable trap geometry.

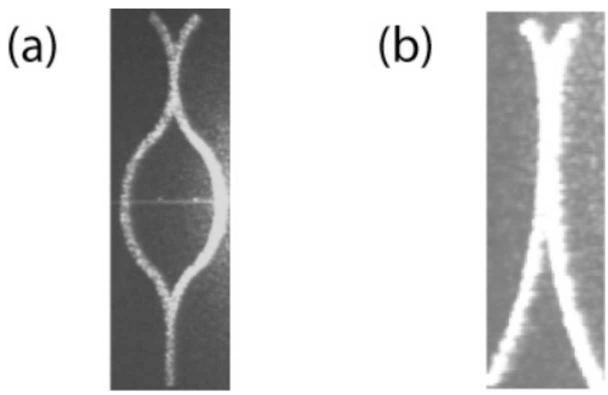


Figure 5: (a) Mach-Zender Interferometer pattern (b) Y-splitter pattern.

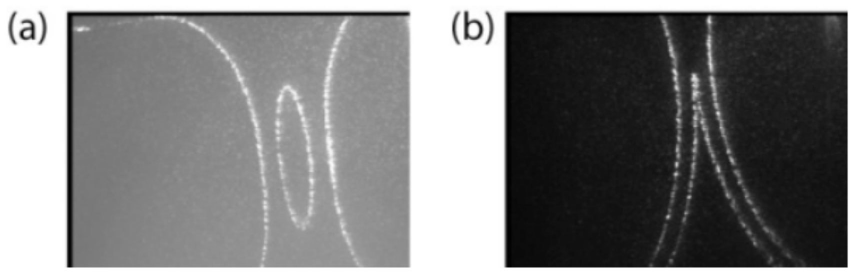


Figure 6: (a) Blue-detuned Mach-Zender Interferometer pattern (b) Blue-detuned Y-splitter pattern

Likewise we can also make a blue detuned Y-splitter. The pattern (shown in Fig. 6(b)) is again designed as a blue-detuned splitter, a design that is difficult solely using hollow beams.

6.3.4 Array of dipole traps

A grid of dipole traps formed with a conventionally microfabricated lens array has recently been observed in a MOT by the Ertmer group. The same group has also recently observed a similar configuration using VCSEL and microlens arrays, which allows the individual addressability of dipole traps. Such traps have applications in quantum computing. The use of SLMs has now begun to be established in colloidal physics, where the creation of arrays of particles is of interest in the study of particle interactions, colloidal transport and nanofabrication. Here we demonstrate an array of spots that can be used in a similar manner to the ones described above. The array produced is shown in Fig. 7. We see one of the problems of using an essentially diffractive optical element to create the light patterns in Fig. 7(a); that of the zero-order diffraction spot. This is due to the fact that we are not using an ideal continuous phase profile when we use an SLM, but a pixellated, or discrete implementation of such a profile. This problem can be addressed in several ways, first the spot can be spatially filtered from the desired image. Secondly the design of the pattern can take the position of the zero-order spot into account with, say the lattice structure such that the zero-order spot does not obscure any the the surrounding spots (Fig. 7(b)), or by moving the pattern offa xis (Fig. 7(c)).

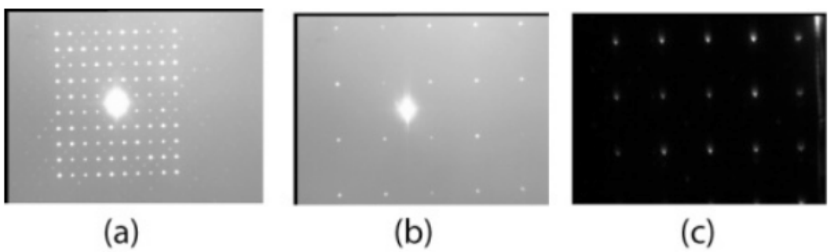
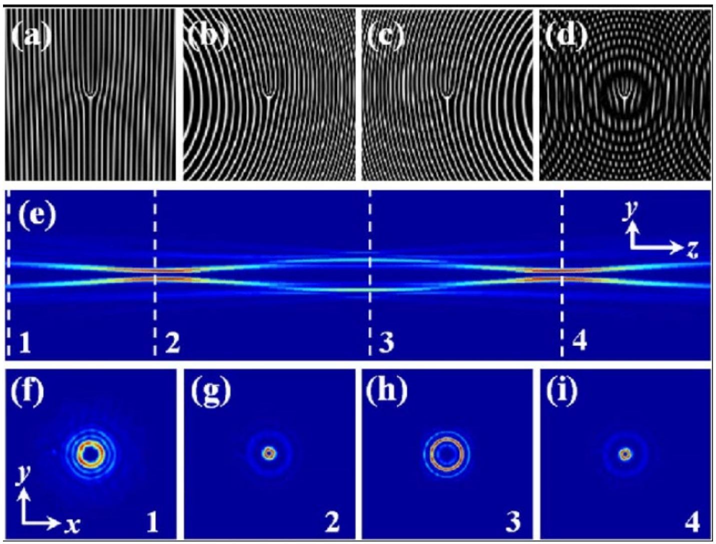


Figure 7: Square array patterns. In (a) we see a ten-by-ten arrays of spots. Here the lattice constant is such that the zeroth order diffraction pattern interferes with the array spots. By increasing the lattice constant we can move the desired pattern away from the unwanted spot. This can then be removed by spatial filtering. Alternatively we can chose to work in a region away from the zeroth order, design the hologram such that the desired pattern is not collinear with the zero order spot, *e.g.* (c) where the zero order spot is seen in the upper right corner.

With the SLM we are able to fully control the relative positions of the trap sites, either between experiments or dynamically, by changing the hologram supplied to the SLM. Hence the lattice constant of the array can be easily tuned, as can the shape of the array. It should also possible to control the position of each trap site in real time, and therefore each site can be turned on or off.

6.3.5 Bottle-beams

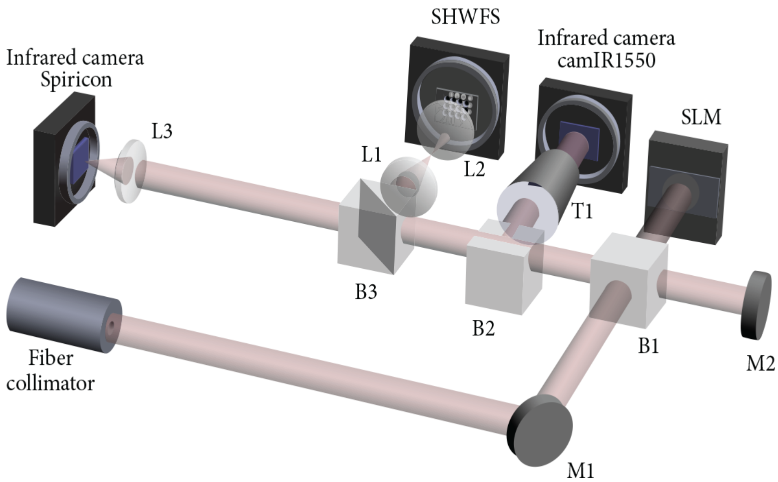
A bottle-beam is a localized dark space region that is surrounded by light. It can therefore be used as a blue-detuned dipole trap. Here we recreate the beam first shown by Arlt and Padgett, which is a superposition of two high-order Laguerre-Gaussian beams with their azimuthal indices l = 0 and their radial mode indices p = 0 and p = 2 respectively.



This dynamical control is readily extended to any pattern we can generate on the SLM, thus we can envisage moving the focus of a beam in the manner of the transport of a BEC with optical tweezers, or generating movable patterns that mimic the optical conveyer belts studied by Kuhr *et al*.

6.4 SPATIAL LIGHT MODULATOR AS AN ATMOSPHERIC TURBULENCE SIMULATOR

The atmosphere is a nonhomogeneous medium that has properties sensitive to changes in pressure, temperature, humidity, and wind speed and direction. Its layered nature and sensitivity to different variables generate turbulent fluctuations in the refractive index of the atmosphere. These turbulent fluctuations are considered to be the principal contributors of the phase fluctuations in traveling optical waves. For decades, scientists have been studying and modeling atmospheric turbulence to obtain a better understanding of its properties and behavior. Atmospheric turbulence simulators (ATS) are powerful tools enabling researchers to evaluate and predict the performance of optical systems prior to field implementation. Some of the most common methods used to simulate atmospheric turbulence are rotating phase plates etched with phase screens, hot plates, or liquid crystal modulators. The use of hot plates is limited by the inability to accurately estimate or precisely control the degree of atmospheric turbulence introduced to the optical path. Rotating phase plates are extremely costly and limited by the number of phase screens. Recent improvements to Liquid Crystal-based Spatial Light Modulators (LCSLM) have made these types of modulators more useful as an ATS for system calibration and design. This is because a given set of aberrations can be carefully introduced and controlled in a laboratory environment for both frozen phase screen studies and real-time turbulence emulations. This paper describes the implementation of such an LCSLM in the short-wave infrared (SWIR) in a laboratory test-bed.



The ATS was designed to enable the evaluation of the effect of atmospheric turbulence on wave propagation in the SWIR regime for FSLC systems and on free-space near-infrared communication links. This simulator innovation is based on a recently patented method for generating atmospheric turbulence. This method produces a time varying phase screen which represents atmospheric turbulence and it is generated by using Karhunen-Loeve polynomials and a splining technique for generating temporal functions of the noise factor for each Zernike polynomial. These generated phase screens are projected on the liquid crystal spatial light modulator. The atmospheric turbulence simulator with the implemented software enables the user to control different parameters such as the aperture of the optical system, the Fried parameter,r_0 and the coherence time of the atmospherer_0. The atmospheric turbulence can be adjusted from weak to strong turbulence conditions by adjusting and controlling the r_0 and/or the optical system aperture. The NRL software algorithm also generates controlled phase aberrations based on Zernike polynomials for system alignment and fine-tuning.

6.4.1 Laboratory Experiment and Testing Procedure

An optical bench was configured in the laboratory to investigate the impact of controlled levels of turbulence and wave front aberrations on a 1550 nm beam. To do this, we used a 1555 nm coherent light source, coated optical elements for the corresponding wavelengths, the SLM to control aberrations, a Shack-Hartman detector, and cameras. The optical configuration is shown in Figure 7. Starting with the light source, we selected a 1555 nm distributed feedback laser (DFB) with a polarization maintaining (PM) fiber. PM fiber components and selected optics were used and were necessary to maintain matching linear polarization through the system and to present a 0° horizontal polarization angle to the LCSLM. These components with careful alignment were required to achieve optimal performance. A commercial collimator with a beam diameter of 6.9 mm and a full-angle divergence of 0.016° was used to launch the light into free space. This diameter was selected to match the 7 mm diameter phase screens generated on the SLM display.

This particular LC SLM is a reflective device. Therefore, we built a Michelson Interferometer using the LC SLM as the variable arm of the interferometer and a gold coated, flat mirror in the reference arm. A Michelson Interferometer reflects light at 0° incidence angle, thereby enabling the use of the entire dynamic range of the device. The beam was then directed through a series of beamsplitters to send portions of the light to four different detection configurations. As shown in Figure 7, the transmissive part of the beam after the first beamsplitter, B1, was sent to the SLM and the reflected portion of the beam was sent to the reference mirror, M2. The light from the reflected beam was directed to (1) a CamIR1550 infrared camera with an aperture-matched telescope, which recorded the interferogram to assess wavefront characteristics at 1550 nm; (2) a Shack Hartman Wavefront Sensor (SHWFS) for wavefront reconstruction; and (3) a Spiricon Infrared Camera to record the Point Spread Function (PSF) of the laser beam also at 1550 nm, without interference from the reference beam. The reference beam was blocked at the time to measure the PSF of the laser beam to avoid interference effects on the measurements.

The two scientific cameras used in the experimental configuration were a CCD sensor from Applied Scintillation Technologies (CamIR1550) and a Silicon CCD from Spiricon Inc. The CamIR1550 is an inexpensive CCD which has 752 (H) × 582 (V) pixels with an effective pixel size of 20 μm. The USB Spiricon L230 is a Silicon CCD which has 1616 (H) × 1216 (V) pixels. We also used a Shack-Hartman Wave front Sensor (SHWFS) consisting of a microlens array of 127 sub-apertures in a hexagonal geometry, a focal length of 18 mm and a 300 μm pitch. FrontSurfer from OKO Technologies was the software implemented for wavefront reconstruction and analysis.

The LC SLM was controlled using the Atmospheric Turbulence Simulator (ATS) application developed by Wilcox at the Naval Research

Laboratory. Modifications to the ATS software were needed to enable integration of the SLM into our system for our application. The NRL ATS software can be customized to control the SLM to produce the phase modulations and simulated turbulence levels through a Graphical User Interface (GUI). The application provides menu-driven options to select the order of aberrations to be generated based on the Zernike polynomials and can be used for fine-tuning the system's optical alignment. The ATS software also provides the option to choose the degree of atmospheric turbulence to be simulated (weak or strong turbulence). This value can be generated by changing parameters like the Fried parameter (r_0) and the system's aperture (D). Different algorithms to generate turbulence can also be used. The algorithms available in the ATS are Zernike Modes, Karhunen-Loeve Modes, and Frozen Seeing.

SUMMARY

- A spatial light modulator (SLM) is an object that imposes some form of spatially varying modulation on a beam of light. A simple example is an overhead projector transparency. Usually when the phrase SLM is used, it means that the transparency can be controlled by a computer.
- Spatial light modulator (SLM) is a general term describing devices that are used to modulate amplitude, phase, or polarization of light waves in space and time. HOLOEYE´s Spatial Light Modulator systems are based on translucent (LCD) or reflective (LCOS) liquid crystal micro displays.
- The use of LC materials in SLMs is based on their optical and electrical anisotropy. A certain gray level represents a defined average voltage across the LC cell. This voltage leads to a variable tilt of the LC molecules due to their electrical anisotropy.
- A spatial light modulator (SLM) is a special device that can manipulate light by modulating the amplitude, phase or polarization of the light waves in the two dimensions of space and time. This means that light is manipulated in order to obtain a desired output, and SLM is commonly used in overhead projectors such as those used in schools and office conference rooms.
- A spatial light modulator is an electronically programmable device that can modulate light output based on a specific fixed spatial pattern (pixel), essentially projecting light that is controlled in either amplitude only, phase only or both (phase-amplitude). This device makes use of liquid crystals to modulate the light, which is why overhead projectors are called LCD projectors.

REFERENCES

1. A. A. Zinchik, Application of spatial light modulators for generation of laser beams with a spiral phase distribution, Sci. Tech. J. Inf. Technol., Mech. Opt. (2015)
2. A. M. Weiner, Femtosecond pulse shaping using spatial light modulators, Rev. Sci. Instrum. 71, 1929-1960 (2000)
3. C. Rosales-Guzmán and Andrew Forbes, How to Shape Light with Spatial Light Modulators, SPIE Spotlight (2017)
4. E.A. Cornell and C.E. Wieman, "Bose-Einstein condensation in a dilute gas, the first 70 years and some recent experiments," Rev. Mod. Phys. **74**875–893 (2002).
5. Freegarde and K. Dholakia, "A cavity-enhanced optical bottle beam as a mechanical amplifier," Phys. Rev. A66 013413 (2002).
6. M.D. Barrett, J.A. Sauer, and M.S. Chapman, "All-Optical Formation of an Atomic Bose-Einstein Condensate," Phys. Rev. Lett.87 010404 (2001).
7. N. Collings et al., The Applications and Technology of Phase-Only Liquid Crystal on Silicon Devices, J. Disp. Technol. 7, 112-119 (2011)
8. N. Schlosser, G. Reymond, I. Protsenko, and P. Grangier, "Sub-poissonian loading of single atoms in a microscopic dipole trap," Nature , 411, 1024–1027 (2001)
9. N.V. Morrow, S.K. Dutta, and G. Raithel, "Feedback control of atomic motion in an optical lattice," Phys. Rev. Lett.88 093003 (2002).
10. O. Houde, D. Kadio, and L. Pruvost, "Cold atom beam splitter realized with two crossing dipole guides," Phys. Rev. Lett. **85**5543–5546 (2000).
11. R. Dumke, T. Muether, M. Volk, W. Ertmer, and G. Birkl, "Interferometer-Type structures for guided atoms," Phys. Rev. Lett.89 220402 (2002).
12. S. Kuhr, W. Alt, D. Schrader, M. Müller, V. Gomer, and D. Mechede, "Deterministic Delivery of a Single Atom," Science **293**278–280 (2001).
13. T. Loftus, C.A. Regal, C. Ticknor, J.L. Bohn, and D.S. Jin, "Resonant Control of Elastic Collisions in an Optically Trapped Fermi Gas of Atoms," Phys. Rev. Lett.88 173201 (2002).

14. T.L. Gustavson, A.P. Chikkatur, A.E. Leanhardt, A. Görlitz, S. Gupta, D.E. Pritchard, and W. Ketterle, "Transport of Bose-Einstein condensates with optical tweezers," Phys. Rev. Lett.88, 020401 (2002)

CHAPTER 7

PHOTOREFRACTIVE CRYSTALS

INTRODUCTION

Photorefraction is an effect of refractive index variation in the material exposed to the light. This property makes it possible to create spatial distribution of refractive index in the crystal and can be used in wide variety of devices including spatial light modulators, optical memory, holography, phase conjugation etc.

The photorefractive effect is a nonlinear optical effect seen in certain crystals and other materials that respond to light by altering their refractive index. The effect can be used to store temporary, erasable holograms and is useful for holographic data storage. It can also be used to create a phase-conjugate mirror or an optical spatial soliton.

7.1 THE PHENOMENON OF PHOTOREFRACTION

A qualitative model of the photorefractive effect discovered by Ashkin et al. is as follows. The photoelectrons excited in the illuminated area of a crystal by incident light leave the region through diffusion or drift due to an externally applied electric field (or via the photovoltaic effect). They are trapped in non-illuminated regions to give rise to a non-uniform space charge and, hence, to an electric field distribution within the crystal. The electric field, in turn, causes a non-uniform distribution of the refractive index through the electro-optic effect. This principle of image recording suggests immediately the main aspects of the physics of photorefractive media:

- The nature of photosensitive centers that absorb light and produce mobile charge carriers, mostly electrons.
- The formation of a non-uniform space charge (diffusion and drift of carriers, relaxation effects, influence of the space-charge field itself on its formation process).
- Electro-optic effects in a non-uniform electric field. Analysis of spatial variations of the refractive index that are directly related to the charge distribution pattern within the crystal.
- Propagation and diffraction of light in the crystals with a spatially non-uniform refractive index.

7.1.1 Photosensitive Centers

Despite the vital importance of charge-carrier photoexcitation and trapping (by shallow and deep traps) for the information recording mechanism in PhotoRefrflCtive Crystals (PRCs), the mechanism is still understood incompletely. Only in certain situations can we speak of generally accepted ideas. It is quite apparent that the important condition for light-induced space charge ·to be formed is the presence of impurity centers in the band gap that serve both as the sources of excited photoelectrons (donors) and as electron trapping centers (deep traps or acceptors). In the simplest, though fairly often encountered case, both the donors and traps are impurities of ions of one and the same type of atom but in different valence states.

Let us consider Fe^{2+} and Fe^{3+} ions as examples. They give rise to impurities that are thought to be most important for information recording in such crystals as $LiNbO_3$, $KNbO_3$, $BaTiO_3$, and others. In oxide crystals Fe ions can be present either because of Fe impurities in the starting material or as a result of doping. Where Fe ions are located is often unknown (whether they substitute for certain cations in the crystal, are interstitial ions, or form other types of defects). For instance, in $LiNbO_3$ and $KNbO_3$, Fe ions substitute for Nb^{5+}, the local electroneutrality being preserved in this case by creation of an oxygen vacance near the Fe^{2+}, that is, an Fe^{2+}-V 0 center is formed. Along with Fe^{2+}, Fe^{3+} is present, too.

Upon illumination of the crystal, the light is absorbed by Fe^{2+} impurities, further ion ionization with the formation of Fe^{3+} +e^- occurs, and the excited electron migrates from the illuminated region until it is captured by a deep trap, for instance another Fe^{3+} ion in the non-illuminated region of the crystal.

Typical energies for photoexcitation of Fe^{2+} ions in $LiNbO_3$ and $KNbO_3$ crystals are 3.1-3.2 eV.

The concentrations of Fe^{3+} and Fe^{2+} ions can vary in a wide range (10^{16} - 10^{19} cm^{-3}) with crystal doping. The proportion between Fe^{2+} and Fe^3+ can also greatly differ depending on the subsequent treatment (annealing in reducing or oxidizing atmosphere). The presence of Fe impurities changes the crystal's conductivity noticeably, and affects the drift and diffusion lengths of electrons. These values vary from tens of Angstroms to tens of micrometers from crystal to crystal.

As mentioned above, the photosensitive centers in the form of Fe impurities are typical of $LiNbO_3$, $KNbO_3$, and $BaTiO_3$. The most important type of photoactive centers in the highly popular and practically significant crystals $Bi_{12}SiO_{20}$, Bi_{12} GeO_{20}, and Bi_{12} TiO_{20} have not yet been determined unambiguously. The models given in the literature assume, e.g., Si (or Ge) vacancies, a complex BiO_7 ion bonded to a silicon or oxygen vacancy, the presence of chromium impurities, and so on.

Not only electrons, but often also holes, can participate in the photoexcited charge transport. However, because of a much lower mobility of holes as compared with that of electrons - and a shorter lifetime - their contribution to photorefraction is typically small. Nonetheless, several researchers have found that holes play an important role in the space-charge formation.

For ferroelectric crystals (or, generally speaking, for polar crystals, including the pyroelectric ones), an essential mechanism that controls the electron transport in the absence of the external field is the photovoltaic effect. The essence of the phenomenon is readily understood if we consider the situation when the probability of the excited-electron motion in one or another direction (impulse direction) is anisotropic, and a preferred electron motion generates a resulting EMF. Necessary conditions for the photovoltaic effect to arise are that the photoactive center itself be spatially asymmetric, that it has a dipole moment, and that such centers be preferentially oriented to prevent the total averaging of the directions of the photoexcited electron transfer. Equally important contributions to the photorefractive effect come from the anisotropy of carrier recombinations, scattering anisotropy, and so on. Such necessary conditions for the photovoltaic effect are encountered in polar crystals. The observed photovoltaic fields in dielectric crystals of the $LiNbO_3$ type can be as high as 10^4 to 10^5 V/cm.

Generally speaking, nonpolar crystals also exhibit the phenomenon of photo-EMF generation. Here we mean the crystals that lack a center of inversion where the linear and circular photovoltaic effects are observed. The preferred direction in space is then governed by the direction of linear or circular light polarization. This generated voltage is, however, typically small, though exceptions are possible, for instance, if the resistivity of the crystal is sufficiently high.

7.1.2 Mechanism for Optical Information Recording

Linear Approximation of Information Recording

The photorefractive crystals offer the capability for recording both images and holograms. For either of these cases we shall use the term “information recording”. Depending on the task to be solved, recording is accomplished using either ordinary incoherent light or laser radiation. For research into the properties of crystals, however, recording of simple sinusoidal gratings formed by interference of two coherent beams is, as a rule, preferred (though not always). The assumption of linearity of information recording in photorefractive crystals is mainly responsible for the wide popularity of such a research technique.

A fact is that no matter how complex the three-dimensional recording intensity pattern I(x, y, z) may be, it can be represented by a superposition of cosinusoidal and sinusoidal patterns of the type $I(K)\cos(K\cdot r)$, $I(K)\times\sin(K\cdot r)$, or, in exponential form, by $I(K)\exp(iK\cdot r)$, where $I(K)$ is the Fourier amplitude of the light intensity, $K\cdot r = 2\pi(\nu x+\xi y+\gamma z)$, and K is the wave vector of the pattern with the projections $K_x=2\pi\nu$, $K_y=2\pi\xi$, and $K_z=2\pi\gamma$. The values ν, ξ, γ are called the spatial frequencies; $\nu=1/\Lambda_x$, $\xi=1/\Lambda_y$, $\gamma=1/\Lambda_z$, where Λ_x, Λ_y, Λ_z are the grating spacing’s in the X, y, and z directions, respectively. The tradition has been established in the literature on photorefractive media to refer to the wave vector projections K_x, K_y, and K_z as spatial frequencies as well; no confusion has arisen so far.

Thus, in the most general case, the light intensity to be recorded can be given by the Fourier integral

$$I(x,y,z) = I_0 \int\int\int m(\nu,\xi,\gamma) exp\left[-i2\pi(\nu x + \xi y + \gamma z)\right] d\nu d\xi d\gamma, \tag{1}$$

where I_0 is the average intensity, and $m(\nu,\xi,\gamma)$ is the relative spectral (spatial) density of the light intensity. In specific cases, expansion in a Fourier series of sines and cosines functions is possible. Then, with the same notations, $m(\nu,\xi,\gamma)$ implies a relative spectral amplitude or coefficient of light-intensity spatial modulation for given values of ν, ξ, and γ.

In a similar fashion, the crystal response, i.e., the spatial distribution of the refractive-index change $\Delta n(x,y,z)$, arising in the crystal, can be expanded in a Fourier series or represented by a Fourier integral.

If we assume that the recording mechanism is linear, in other words, the principle of superposition is applicable, recording of each of the sinusoidal patterns making up the complex image will occur, regardless of the presence of other sinusoidal gratings, with one sinusoidal light-intensity pattern producing only one refractive-index grating. Consequently, to understand properly the crystal parameters which determine the recording of complex patterns, it suffices to know how the recording of individual sinusoidal patterns that differ from one another by the spatial frequency alone, proceeds. Sinusoidal patterns can readily be produced by interfering two coherent plane waves. Therefore we shall quite justifiably confine the analysis of all recording processes in the following discussion to that of simple sinusoidal gratings, with the aim of establishing the dependence of the amplitude and phase of the recorded refractive-index grating on spatial frequencies.

An alternative approach, entirely equivalent to that discussed above, in the framework of linear space-invariant systems theory is in general possible, too. It involves the investigation of the crystal response to an input delta function, i.e. the impulse response. From the standpoint of linear systems theory, the two methods are formally equivalent because they are related by the Fourier transformation. Experimentally, however, the study of light diffraction from the refractive index grating presents a more convenient research tool than the tiresome analysis of details of the impulse response shape.

Therefore, in what follows, we perform the analysis in terms of "elementary gratings", with the term "grating" used to describe the sinusoidal distribution of charge, electric field, refractive index, etc. We emphasize that, though the linear approximation is a very useful research tool, it is not always applicable to photorefractive crystals.

Diffusion and Drift Models

We tum now to information recording mechanisms in photorefractive crystals and restrict ourselves, in this section, to the formation of a non-uniform space charge and electric field.

There exist two basic recording mechanisms, namely, diffusion and drift. In the diffusion model, the electrons excited by light migrate from the illuminated region, where their concentration is higher, toward the non-illuminated regions, where the carrier concentration is lower. There they become trapped. Let the crystal be exposed to an interference pattern of the type

$$I(x) = I_0\left[1 + mcos\left(K_x x\right)\right]. \qquad (2)$$

The light intensity is uniform along the y and z axes. In the diffusion model, the redistribution of the photoexcited electrons gives rise to three charge gratings (Fig. 1). The first is the grating of positively charged donors, with the charge density given by

$$\rho^+(x) = \rho cos\left(K_x x\right) + \rho_0, \qquad (3)$$

where

$$\rho = emI_0 t_{ex}\, \beta\alpha / (\hbar w).$$

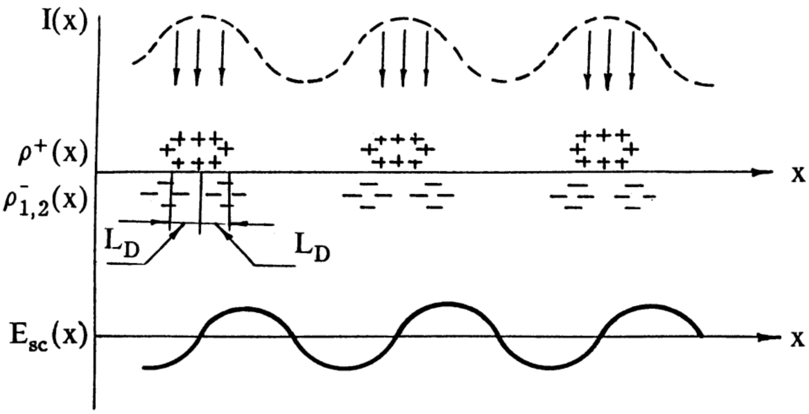


Figure 1. Formation of charge and electric-field gratings through diffusion on illumination by the recording light.

Here, ϱ_0 is the mean density of positive charge, α is the absorption coefficient, e is the electronic charge, t_{ex} is the exposure time, β is the quantum efficiency of electron excitation, and ħw is the energy per recording light photon. The two other gratings are the gratings of negatively charged traps with the charge density for short exposures (the initial stage of recording), i.e.,

$$\rho_1^-(x) = -\frac{1}{2}\rho cos\left[K_x\left(x+L_d\right)\right] - \frac{1}{2}\rho_0$$

and

$$\rho_2^-(x) = -\frac{1}{2}\rho cos\left[K_x\left(x-L_d\right)\right] - \frac{1}{2}\rho_0. \tag{4}$$

These gratings arise because of electron diffusion to the right and left from the site of the excitation. Here, L_d is the diffusion length, i.e., a typical mean distance the electron moves from the excitation site to the trapping location. According to Poisson's equation

$$divE(x) = \frac{\rho(x)}{\epsilon\epsilon_0}, \tag{5}$$

where $\rho(x) \aleph \rho^{\aleph}(x) \quad \rho_1(x) \quad \rho_2(x),$ the resultant field $E_{sc}(x)$ formed by the three gratings is

$$E_{sc}(x) = \frac{\rho}{\epsilon\epsilon_0 K_x}\left[1-cos\left(K_x L_d\right)\right]sin\left(K_x x\right), \tag{6}$$

where ϵ_0 is the free-space electric permeability, and ϵ is the relative dielectric constant of the crystal. The grating field is along the x axis. Other components, along the y or z axes, are zero. For a short diffusion length L_d such that $K_x L_d \ll l$ (this situation is fairly common in photorefractive media),

$$E_{sc}(x) = \frac{\rho L_d^{\,2}}{2\epsilon\epsilon_0} K_x sin\left(K_x x\right). \tag{7}$$

Note that the field grating is shifted by $\pi/2$ with respect to the interference pattern (cos(K_xx) has been transformed into sin(K_xx)] and the grating amplitude is proportional to the spatial frequency K_x. Moreover, comparison of (6 and 3) also reveals that the buildup rate of the field grating grows with increasing L_d and decreasing ◎, with the recording light intensity being the same.

The other limiting case, when $L_dK_x \gg 1$, needs a more thorough treatment.

As the exposure grows, the grating field increases and begins to inhibit diffusion of electrons. At last, the amplitudes of the charge and field gratings grow no longer and steady state is reached; in other words, despite the presence of light that excites electrons, no further growth of the grating field amplitude occurs. This situation arises when the grating field fully compensates for the so-called effective diffusion field E_D. The fact is that diffusion of electrons from the regions of higher concentration to those of lower concentration, which occurs exclusively at the expense of thermal motion, can be represented as electron motion in an effective electric field whose magnitude is

$$\pm_D(\ )=\frac{k_BT}{e}\frac{1}{n(x)}\frac{\pm(\ )}{dx} \qquad (8)$$

where k_B is the Boltzmann constant, Tis the temperature, n(x) is the electron concentration at point x, and dn(x)/dx is the concentration gradient. The grating field that is a real electric field is opposite to $E_D(x)$. Steady state is reached when $\left|E_{\pm}(x)\right|=\left|E\ \ (x)\right|$. The grating field E_{sc} at steady state is independent of such crystal parameters as the dielectric permeability or the diffusion length.

Since $n(x)\propto I(x), E_D(x)$ is shifted by $\pi/2$ with respect to the interference pattern and is proportional to the spatial frequency. The diffusion field for a sinusoidal grating with $m \ll 1$ can be written as

$$E_D(x)=mE_D sin(K_x x), \qquad (9)$$

where

$$E_D=K_x\frac{k_BT}{e}. \qquad (10)$$

The condition $m \ll 1$ implies that there is a fairly strong uniform illumination of the crystal and, hence, a high level of photoconductivity. In further discussion of the diffusion mechanism we shall refer to the amplitude E_D as the diffusion field. For T = 300K and $K_x/2\pi \simeq 10^2$ - 10^3 mm^{-1}, E_D is 0.1-1 kV/cm.

We now tum to the drift recording mechanism in which, as distinguished from the diffusion mechanism, the motion of the photoexcited charge occurs in an external electric field (Fig. 2). The photoexcited electrons move in one direction and pass, on the average, a certain typical distance L_0 before being trapped. Here L_0 denotes the drift length and is related to the external field E_0 by $L_0=\mu\tau E_0$, μ being the mobility and τ the lifetime of a photoexcited electron.

The essential features of the drift mechanism are readily understood from the following model. We assume, as before, that the crystal is illuminated by light with $I(x) = I_0[l+m\cos(K_x x)]$. The external field is along the x axis. In the initial stage of recording, when the electron transport is still unaffected by the grating field, the charge distribution for short drift lengths can be represented as a sum of two gratings p+(x) = pcos(Kxx)+p0 and $\varrho^-(x) = - \varrho\cos[K_x(x+L_0)]-\varrho_0$, which are formed by positively charged donors and negatively charged traps; ϱ is determined from (3). The net charge density is

$$\rho(x)=\rho\Big(cos(K_x x)-cos\big[K_x(x+L_0)\big]\Big). \qquad (11)$$

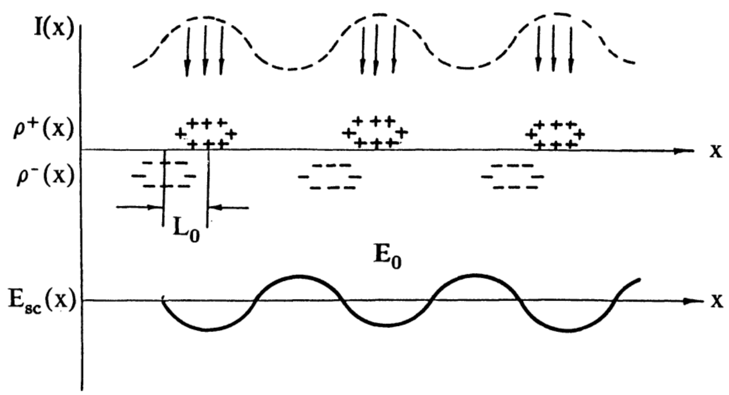


Figure 2. Formation of charge and electric-field gratings in a photorefractive crystal through drift.

For short drift lengths K_x L_0 ≪ l, the field of such a grating is, according to (5),

$$E_{sc}(x)=\frac{-\rho}{\epsilon\epsilon_0}L_0\,cos\left[K_x\left(x+\frac{1}{2}L_0\right)\right]. \qquad (12)$$

As seen from (12), in this case the grating field is almost in phase (unshifted grating) with the recording light grating, and its amplitude is independent of K_x. but depends on the drift length. In the other limiting case (L_0K_x ≫ l), a negative charge grating is not formed since electrons are uniformly "spread" throughout the crystal. Only the grating $\varrho^{+}(x)+\varrho_0$ is left, which, according to (5), gives the field grating shifted by π/2 with respect to the light intensity grating, and the field amplitude is $\propto 1/K_x$.

For long exposures, when the recording exhibits a steady-state behavior, the grating-field amplitude attains its highest possible value. More sophisticated consideration is needed to analyze how the grating reaches this regime. Here we note only that in the drift model the maximum grating field is limited either by the external field E_0, or (as in the diffusion mechanism) by depletion of donors and traps (the typical field value is here denoted by E_q). We define E_q as the maximum amplitude of a sinusioidal grating field when depletion of either donors or traps occurs. Thus, the field E_q is limited by the lower of the two values, i.e., the donor concentration N_D or the trap concentration N_A. If we denote this lower concentration by N_{min}, then

$$E_q = \frac{eN_{min}}{\epsilon\epsilon_0 K_x}. \tag{13}$$

Since E_0 and E_q can be well above E_D for moderate K_x, the efficiency of the drift mechanism can be appreciably higher.

The drift mechanism provides a great variety of recording conditions. The external field can be applied perpendicular to the grating fringes or, alternately, an AC or DC field can be used, the crystal's electrode contacts may be ohmic or rectifying, the crystal can be insulated from electrodes by a dielectric layer, etc .. All these factors affect the recording efficiency.

Of utmost practical significance is the recording configuration in which the electric field is applied along the z axis and the recording is performed by light incident on the crystal surface that is normal to the z axis. This geometry is commonly employed for recording images in spatial light modulators. In this case, a grating $\varrho(x)$ having a finite thickness along the z axis is formed. In the linear mode of operation (as long as the field of the grating itself does not affect the drift of electrons) the grating thickness in practically important cases is about the same as the drift length L_0. Depending on the particular conditions of the current through the sample upon illumination, either the negative or positive charge grating may dominate the process. Because of the finite thickness of the charge grating, nonzero values of the field are observed not only along the x axis, but also along the z axis. The field grating $E_x(x)$ is always shifted by π/2 from the interference pattern when the external field is along the z axis, and the grating $E_z(x)$ is in phase with it. The spatial frequency dependences of $E_x(x)$ and $E_z(x)$ also turn out to be different.

The drift mechanism can take place not only in an external electric field, but also because of the photovoltaic effect mentioned above. From the formal point of view, the electron transport caused by the photovoltaic EMF for $m \ll 1$ can be represented as electron drift in an effective external field. The drift mechanism in information recording by means of the photovoltaic effects is therefore nearly equivalent to the process taking place in an external field.

7.1.3 Electro-Optic Effects

An electro-optic effect is a change in the optical properties of a material in response to an electric field that varies slowly compared with the frequency of light. The term encompasses a number of distinct phenomena, which can be subdivided into changes in absorption and changes in refractive index.

Longitudinal and Transverse Electro-Optic Effects

A non-uniform charge - and thus electric field - distribution induced in a photorefractive crystal cause variation of the index of refraction through the electro-optic effect.

In this section, electro-optic phenomena will be discussed from a general point of view, with particular attention given to specific features for photorefractive crystals.

The PRCs may originally be either isotropic or anisotropic. In an isotropic crystal, the dielectric permeability ϵ^w is a scalar. The refractive index ($n = \sqrt{\epsilon^w}$) and the light-propagation velocity depend neither on the direction of light-wave propagation nor on the direction and state of its polarization. Optical activity is ignored here. In an anisotropic crystal, $\hat{\epsilon}^w$ is a tensor, and the refractive index will generally depend on both the propagation direction and polarization of this wave. Nonetheless, one or two directions can exist in a spatially homogeneous crystal, the refractive index of the light traveling along which is independent of the wave polarization. These directions are called optical axes.

The electro-optic effect consists of a change of the refractive index of the crystal and a change of the orientation of its optical axes on application of an external electric field. In terms of crystal optics, the electro-optic effeet rests on the dependence of components of the dielectric impermeability tensor $\alpha_{ij}^{\omega}, \hat{\alpha}^{\omega} = \left(\hat{\epsilon}^{w}\right)^{-1}$, on the electric field E. If α_{ij}^{ω} and E are linearly related, the effect is termed the linear electro-optic (Pockels) effect. If the relation is quadratic, we have the Kerr effect. In this monograph we deal exclusively with the Pockels effect.

A necessary "existence" condition for the Pockets effect is the absence of an inversion center in the crystal. The linear relation between α_{ij}^{ω} and E is typically given by

$$\Delta\alpha_{ij}^{\omega} = \sum_{n} r_{ijn} E_n = r_{ijn} E_n, \tag{14}$$

where E_n is the electric-field projection (n = x, y, z), and the third-rank tensor r_{ijn} is the electro-optic tensor. Since $\Delta\alpha_{ij}^{\omega} / \alpha^{\omega}{}_{ij} \ll 1$, the component of the dielectric permeability tensor for a cubic crystal acquires the form

$$\epsilon_{ij}^{\omega}\left(E\right) \simeq n^2 \delta_{ij} - n^4 r_{ijn} E_n, \tag{15}$$

where $\delta_{ij} = 1$ for i = j, and $\delta_{ij} = 0$ for i ≠ j. Equation (15) reveals that, in the most general case, the orientation of the coordinate system where tensor $\hat{\epsilon}^{\omega}$ is diagonal, will depend on the electric-field E direction, and the refractive indices in this coordinate system $\left(n_1 = \sqrt{\epsilon_{11}^{\omega}},\ n_2 = \sqrt{\epsilon_{22}},\ n_3 = \sqrt{\epsilon_{33}}\right)$ depend on the magnitude of the field. When field E is applied, the originally optically isotropic crystal ($n_1 = n_2 = n_3$) can become uniaxial ($n_1 = n_2 \neq n_3$) or biaxial ($n_1 \neq n_2 \neq n_3$). Let us consider some features of the electro-optic effect, taking the originally uniaxial and isotropic crystals as illustrations. A plane light wave with an arbitrary propagation direction and an arbitrary linear polarization direction in the crystal can be represented as a superposition of two so-called eigenmodes. These modes are identified as waves with polarizations oriented perpendicular to each other and traveling through the crystal with their own refractive indices. One of the eigenmodes is the wave with the polarization normal to the optical axis and the direction of propagation. This wave is referred to as the "ordinary" wave, and the "ordinary" refractive index n_0 corresponds to it. Once the ordinary wave is defined, the second mode is determined unambiguously and is called the "extraordinary" wave, with the corresponding "extraordinary" refractive index n_e. Note that n_0 is the same for all ordinary waves in the crystal, and n_e depends on the propagation direction of the extraordinary wave. For an arbitrarily incident wave, the ordinary and extraordinary beams will refract at different angles - because $n_0 \neq n_e$ - to give rise to birefringence.

If the original wave is incident perpendicular to the crystal surface and the optical axis is parallel to the crystal surface, the ordinary and extraordinary waves travel in the same direction within the crystal, but, because of different propagation velocities (generally n_0 is not equal to n_e), the waves have different phase delays. The difference between phase delays is

$$\phi_{oe} = \frac{2\pi d\left(n_0 - n_e\right)}{\lambda}, \qquad (16)$$

where d is the crystal thickness. Application of the electric field can cause variations in n_0 and n_e, and a change of φ_{oe}, i.e., $\Delta\varphi_{oe}$ arises. Variation in phase retardation induced by the applied field can easily be demonstrated experimentally. This is precisely the effect used in polarization techniques for studying electro-optic properties of crystals, and also in the design of electro-optic light modulators.

The electro-optic effects can be classified depending on the orientation of the applied field E and the light-propagation direction. If E||K, the longitudinal effect is possible; if E⊥K, the transverse effect can arise (Fig. 3). Apparently the values of $\Delta\varphi_{oe}$ for both the longitudinal and transverse effects will depend on how the crystal is cut, in other words, on the direction in which the light travels with respect to the crystallographic axes and electric-field vector E.

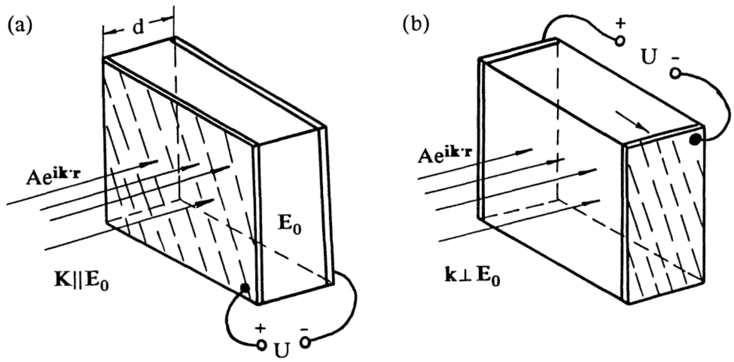


Figure 3. Mutual orientations of the applied electric field E_0 and light propagation direction (direction of K.) for (a) the longitudinal electro-optic effect, and (b) the transverse electro-optic effect.

The simplest relations are obtainable for an originally isotropic (cubic) crystal. With n_0 applied field, $n_0 = n_e = n$ and $\varphi_{oe} = 0$. If the field E_ℓ is applied along one of the cubic crystallographic axes- for instance, the [001] direction - the crystal becomes a biaxial one with $n_z = n$, $n_{x'} = n+\Delta n$, $n_{y'} = n-\Delta n$, where $\Delta n = ½n^3\ r_{41}\ E_\ell d$. We assume here that the z axis || [001], and the two other axes x′ and y′ are at ±45° to the (100] axis. For light travelling along the z axis (the longitudinal effect), the eigenmodes are the waves polarized along x′ and y′. Because calculations of phase relations for these modes are similar to those for ordinary and extraordinary waves in uniaxial crystals, we retain here the notation $\Delta\varphi_{oe}$ and, to make it clear, assume that the wave which is polarized along x′ is similar to the ordinary beam, and that polarized along y′ is similar to the extraordinary beam. Then for the longitudinal electro-optic effect

$$\Delta\phi_{oe} = \frac{2\pi d n^3 r_{41} E_\ell}{\lambda} = \frac{2\pi n^3 r_{41} U}{\lambda}, \qquad (17)$$

where U is the voltage applied to the crystal plate. r_{41} is the respective electro-optic coefficient with the subscript 41 being used instead of ijn. when ijn = 321 = 231.

A specific characteristic of the magnitude of the electro-optic effect is the so-called half-wave voltage $U_{\lambda/2}$. It is defined as the value of the applied voltage U at which $\Delta\varphi_{oe}$ becomes equal to π. The half-wave voltage is given by

$$U_{\lambda/2} = \frac{\lambda}{2n^3 r_{41}}. \qquad (18)$$

The half-wave voltage differs for different crystal cuts. However, handbooks typically specify $U_{\lambda/2}$ for such a cut where the half-wave voltage is at the minimum. When $\Delta\varphi_{oe} = \pi$ is reached, the polarization of the resultant wave (superposition of the ordinary and extraordinary beams or corresponding eigenmodes) changes by 90° at the crystal exit (the direction of the input light polarization is shown in Fig. 4). Then the crystal is equivalent to the so-called half-wave plate. The vertically polarized input beam will emerge from the crystal with a horizontal polarization. For an arbitrary $\Delta\varphi_{oe}$, the output polarization is elliptic.

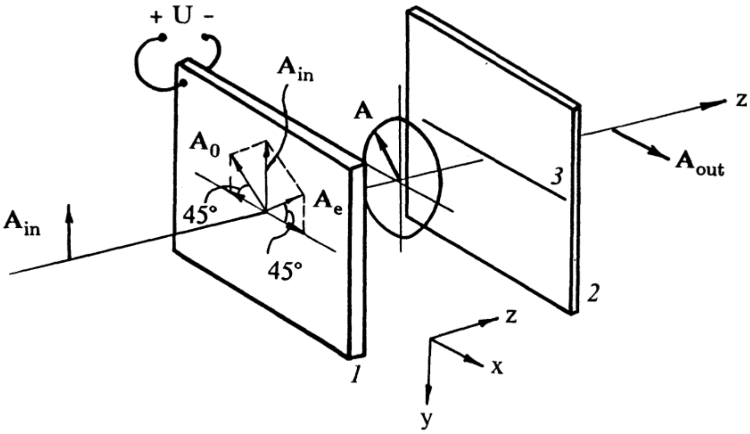


Figure 4. Propagation of linearly polarized light through an electro-optic crystal: (1) crystal, (2) analyzer. (3) orientation of the analyzer axis. A_{in} denotes the incident light with vertical linear polarimtion; A_0 and A_e the ordinary and extraordinary components into which the incident light is decomposed; A is the elliptically polarized beam emerging from the crystal, and A_{out} the horizontally polarized light wave resulting from the elliptically polarized wave passing through the analyzer.

If the transverse electro-optic effect is employed, then, for a cubic crystal,

$$\Delta\phi_{oe} = \frac{2\pi d n^3 r_{41} E_t p}{\lambda}, \qquad (19)$$

where E_t is the field applied in the plane perpendicular to the light-propagation direction, and p is a coefficient ($p \leq 1$) that depends on the crystal orientation. Now the effect is governed rather by the magnitude of the transverse field than by the potential difference, as with the longitudinal effect.

A change in the light-polarization state at the crystal exit, resulting from different phase delays experienced by the ordinary and extraordinary beams, can be converted into the light intensity variations by a polarization analyzer. For instance (Fig. 4), if light is polarized vertically at the crystal input, and an analyzer with a horizontally directed transmission axis is placed behind the crystal, then, in a general case for an arbitrary $\Delta\varphi_{oe}$, the light intensity behind the analyzer ($I_{\rightarrow}$) is given by

$$I_{\rightarrow}\left(\Delta\phi_{oe}\right) = I_{\uparrow} \sin^2\left(\frac{1}{2}\Delta\phi_{oe}\right), \qquad (20)$$

where $I_{\uparrow}$ is the light intensity at the input.

7.1.4 Electro-Optic Effects in PRCs

The notions of the transverse and longitudinal electro-optic effects, and also that of the half-wave voltage, are widely used in polarization techniques for investigating electro-optic media and also for analyzing electro-optic light modulators. Generally speaking, these notions are used in studies of photorefractive crystals as well, however here they acquire somewhat different meanings. These differences are attributable first to the fact that the electro-optic effect arises not in externally applied fields, but in internal fields induced by an exposure of the crystal to recording light. Secondly, we have to deal with a nonuniform (both in magnitude and direction) field within the crystal. Several examples follow.

- The notions of the longitudinal and transverse effects are traditionally associated with the mutual orientation of the applied electric field and the light-propagation direction. In photorefractive crystals, however, the electro-optic effect of interest arises because of internal fields that are caused by a nonuniform space charge, i.e., field E_{sc}. Thus, the notions of the longitudinal and transverse effects are now used to describe the orientation of the light-propagation direction with respect to the space-charge field direction within the crystal. The role of the external field may be unimportant here, since it serves merely to induce the drift of charge carriers. For the diffusion mechanism, the external field may not be applied at all.
- In photorefractive crystals, especially when they are employed in coherent optical systems, extensive use is made of the phenomenon of light diffraction from the refractive index grating. In this diffraction experiment the index variations should be treated separately for the ordinary and extraordinary beams, since light can diffract in an entirely independent manner for the two beams. Of primary concern here is a change of each refractive index n_0 and ne by itself, not of their difference. For instance, sometimes n_0 and n_e exhibit similar dependences on the electric field, but induce no changes in birefringence ($\Delta\varphi_{oe} = 0$), and are not detected by the polarization technique, whereas the diffraction experiments allow measurement of the individual variations (n_0 and n_e). Introducing here also the notion of the half-wave voltage is sometimes convenient, but its meaning and magnitude prove to be somewhat different from (18). The half-wave voltage can be specified in this case as the voltage under which the phase retardation for an individual beam alters by π. This parameter is therefore different for different beams (ordinary and extraordinary)

and its magnitude typically differs from that in the definition of $U_{\lambda/2}$ given previously.

We should comment specifically on the use of the refractive index as a performance parameter related to recording of a two- or three-dimensional light intensity distribution I(x, y, z). Take recording of a two-dimensional image as an example. In particular, difficulties are encountered for the transverse electro-optic effect. The fact is that on recording a two-dimensional pattern I(x, y) the non-uniform internal electric field within the crystal changes from point to point not only in magnitude, but also in direction. However, the direction of the transverse electric field in a given region of the crystal is what determines a local coordinate system x′y′ in this region (x′, y′ are the directions of the optical axes). A change in the transverse field direction (in the xy plane) leads to a change in the local coordinate system (x′y′) that establishes the polarization direction of the ordinary and extraordinary beams. This consideration raises doubts as to whether it is reasonable to analyze information recording in PRCs using the description in terms of the refractive index variations, because decomposition of the recording light intensity or the space-charge electric field into spatial frequencies is carried out in a fixed coordinate system, and the refractive indices require a local coordinate system that depends on the field direction and, what is more, changes from one image to another.

In this case, the convention of "eigenmodes" (normal modes) of light waves and refractive indices are meaningless for the crystal as a whole. Thus, to find the spatial two-dimensional dependence of the output light amplitude $A_{out}(x, y)$, we should calculate the light wave field at the crystal output using a more general approach. The general solution of this problem for thin plates. The input-output relation of the complex light amplitude in a thin plate is determined by a tensor that is a linear function of the recording intensity I(x, y). No difficulties arise for the longitudinal electro-optic effect, since the field direction $E_z(x,y)$ is uniquely specified for the entire crystal plane. The situation is also simplified for the transverse effect if spatial frequencies differ markedly in magnitude (either $\nu/\xi \ll 1$ or $\nu/\xi \gg 1$), and also if the initial birefringence is fairly large $(n_0 - n_e) \gg \Delta n(x,y)$. In these cases, a description in terms of a non-uniform distribution of the refractive index $\Delta n(x, y)$ can often be used quite justifiably.

The analysis of the electro-optic effect associated with a non-uniform field in PRCs can give birth to a new branch of the optics of solids, i.e. to electro-optics in non-uniform fields.

7.2 SALIENT FEATURES OF PHOTOREFRACTIVE CRYSTALS FOR HOLOGRAPHY

The mechanism of information recording determines such characteristics of the photorefractive crystals as diffraction efficiency, information capacity, spatial-frequency bandwidth, sensitivity, speed of operation, dynamic range, and storage time. The parameters listed are essential for the description of the media in terms of information theory, but they do not yield adequate estimates for other important features such as reliability, thermal stability, ease of fabrication, and the cost.

7.2.1 Diffraction Efficiency

Evaluating the diffraction efficiency involves calculation of the amplitude of refractive index variation Δn. Let us restrict our analysis to one eigenmode of the light wave in the crystal. Then Δn denotes the refractive index deviation for a given mode (for instance, ordinary or extraordinary beam). Because Δn is unambiguously related to the grating field amplitude through the electro-optic coefficient, the field in the crystal must be determined for a specified charge distribution.

In the linear regime of recording, i.e., when the field of the grating itself still does not affect the charge formation, the complex charge grating can be represented as a superposition of elementary gratings of the positive and negative charges linearly related to the intensity of the recording light. Now the field of each elementary grating can be found separately, and the resultant field is a sum of all fields of the different gratings. This approach reduces the problem to an analysis of the electrostatic factors for a single grating. The linear regime is typical of spatial light modulators and the initial stages of holographic recording (far from the steady-state regime). The amplitude of the elementary-charge grating is proportional to the recording light intensity, and is independent of the spatial frequency, though the amplitude of the resultant grating consisting of a number of elementary gratings can be frequency dependent. For simplicity, we shall use the model of an infinite, originally homogeneous medium, with the term "infinite" implying that the sample sizes in any direction are far greater than the grating spacing. Let us consider three examples: a volume grating, a thin grating, and a thin grating in a sample coated with electrodes.

Volume Grating

First we consider a volume grating with the charge density ϱ, i.e.,

$$\rho(r) = \rho cosKr \tag{21}$$

According to Poisson's equation (5) the grating field is given by

$$E(r) = \frac{K}{K^2}\frac{\rho}{\epsilon\epsilon_0}\sin Kr. \tag{22}$$

It is common to consider the situation when the grating wave vector is along a selected axis of the coordinate system chosen. Take the grating directed along the x axis as an example, i.e., $K = K_x = 2\pi\nu$. Then, instead of (22), we have

$$E_x(x) = \frac{\rho}{2\pi\epsilon\epsilon_0}\frac{1}{\nu}\sin(2\pi\nu x),\ and\ E_y = E_z = 0. \tag{23}$$

Equation (23) predicts that if the light propagates along the z or y axis, only the transverse electro-optic effect must be used; and if the light travels along the x axis, the longitudinal effect must be employed. The field grating and, hence, the refractive index grating are shifted by a quarter of the spacing with respect to the charge grating.

For a cubic crystal of the 43m and 23 point group with the z axis oriented along the [110], with the y axis along the [110], and the x axis along the [001] crystallographic axes, respectively, the transverse effect takes place for light polarized along the y axis. We then have

$$\Delta n(x) = \frac{1}{2}n^3 r_{41} E_x(x) = \Delta n \sin(2\pi\nu x). \tag{24}$$

We assume that the hologram thickness, though large, is finite and equals d. Then, under the condition that $\pi\Delta nd/\lambda\cos\theta_B \ll 1$, Eq. (24) yield for the diffraction efficiency

$$\eta = \left(\frac{1}{4}n^3\frac{r_{41}}{\epsilon\epsilon_0}\frac{d}{\cos\theta_B}\frac{\rho}{\lambda}\frac{1}{\nu}\right)^2. \tag{25}$$

Attention should be paid to the ν^{-2} frequency dependence of the diffraction efficiency, which is fundamental, for a single grating with K along a specified direction.

In a more general case, when the charge grating is arbitrarily oriented, i.e., $K \neq K_x$, the diffraction efficiency is derived in a more complicated manner, and the frequency dependence can be different from ν^{-2}. To avoid misunderstanding, we emphasize that the frequency dependence of η for an arbitrary grating orientation will imply variations of η as a function of the individual spatial frequencies $\nu = K_x/2\pi$, $\xi = K_y/2\pi$, $\gamma = K_z/2\pi$, rather than of the overall modulus K - i.e., η = η (ν,ξ,γ) - because the recorded image I(x, y, z) is decomposed into harmonics with the frequencies ν,ξ,γ.

The complicated behavior of η as a function of spatial frequencies is attributable to the fact that the changes in the dielectric permeability or refractive index are governed by a superposition $\Sigma_j\, r_{inj} E_j$ of the field projections E_j rather than by the magnitude of modulus E. The different projections E_j have different frequency dependences, for instance

$$E_x\left(x\right)=\frac{\rho 2\pi\nu\sin Kr}{\epsilon\epsilon_0 K^2}=\frac{\rho\nu\sin Kr}{2\pi\epsilon\epsilon_0\left(\nu^2+\xi^2+\gamma^2\right)}, \tag{26}$$

while $E_y \propto \xi/K^2$ and $E_z \propto \gamma/K^2$.

However, the condition $\Delta K_{max}/K_c \ll 1$ is typically satisfied in recording complex volume holograms (but not for SLMs). Here we assume that the individual gratings labeled j, constituting the complex hologram, have the frequency $K^j = K_c + \Delta K^j$, where K_c is the average (carrier) frequency, and ΔK_{max} is the maximum deviation of the spatial frequency from K_c. Therefore, the $\eta \propto 1/K^2$ dependence is a dominating one. Note that for the same reason ($\Delta K_{max}/K_c \ll 1$) recording of volume holograms can be adequately· analyzed through the diffraction of eigenmodes and the refractive-index modulation.

Thin Grating

Now we consider examples for the frequency dependence of the diffraction efficiency for thin holograms or SLMs. That the hologram is thin means in this case that the charge grating thickness $h \ll \nu^{-1}$, but the grating spacing ν^{-1}, in tum, is far smaller than the crystal thickness d.

The first example is a thin charge grating in a volume crystal in the z = 0 plane (Fig. 5), namely

$$\rho\left(x\right) = \rho\cos 2\pi\nu x.$$

Under the conditions mentioned above, we can introduce the surface charge density σ = ϱh, and the grating field is then given by

$$E_x(x,z) = \frac{\sigma}{2\epsilon\epsilon_0} e^{-}|2\pi\nu z| \sin 2\pi\nu x = E_x(z)\sin(2\pi\nu x),$$

$$E_y = 0,$$

$$E_z(x,z) = \pm\frac{\sigma}{2\epsilon\epsilon_0} e^{-}|2\pi\nu z| \cos 2\pi\nu x = E_z(z)\cos(2\pi\nu x). \tag{27}$$

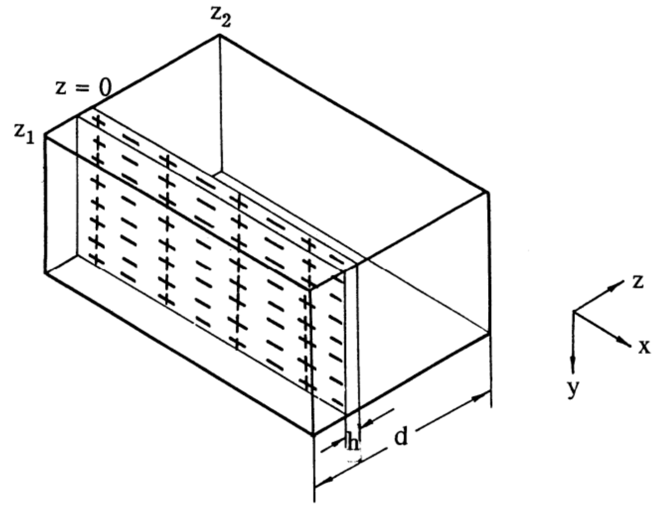


Figure 5. Thin charge grating in a crystal of thickness d (grating thickness: h≪d).

The sign + or - in the last expresion corresponds to z > 0 or z < 0, respectively.

For the transverse electro-optic effect where Δn arises from the field component E_x, the refractive index grating is shifted by π/2 from the charge grating. The amplitude modulation of the readout-beam phase taking account of (24) is

$$\psi_1 = \frac{\pi}{\lambda} n^3 r_{41} \int_{-\infty}^{+\infty} E_x(z)\,dz = \frac{1}{2\lambda} n^3 \frac{r_{41}}{\epsilon\epsilon_0}\frac{\sigma}{\nu}. \tag{28}$$

Since we are dealing with a thin hologram

$$\eta = \left(\frac{n^3 r_{41}\sigma}{4\lambda\epsilon\epsilon_0\nu}\right)^2. \tag{29}$$

Equation (27) reveals that due to the transverse effect the field grating is shifted with respect to the charge grating, and the diffraction efficiency $\eta \propto \nu^{-2}$ and is independent of the crystal thickness.

For the longitudinal effect in this geometry (i.e., when the grating is within the crystal volume), the field Ez has different signs according to z > 0 and z < 0. Therefore, $\psi_1 \propto \int E_z dz = 0$, and $\eta = 0$. If the charge grating is on the crystal surface, $\psi_1 \neq 0$ for the longitudinal effect. Solution of this problem for the longitudinal and transverse effects requires that the corresponding boundary conditions be taken into account.

Thin Grating with Electrodes

We consider one more example of a thin grating of particular concern for SLMs, i.e., a thin charge grating (Fig. 6)

$$\rho(x) = \rho \cos(2\pi\nu x)$$

located in the center of a crystal plate of thickness d with the front and back faces coated with electrodes. The charge-grating thickness is $h \ll \nu^{-1}$; the z axis is normal to the plate plane.

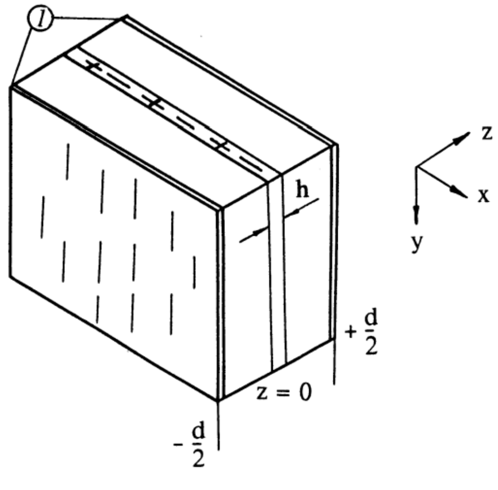


Figure 6. Thin charge grating in the center of a crystal. The front and back faces of the crystal are coated with electrodes.

Then denoting $\varrho h = \sigma$ we have

$$E_x(x,z) = \frac{\sigma\nu}{\pi\epsilon\epsilon_0}\frac{1}{d}\sin 2\pi\nu x \sum_{j=0}^{\infty} \frac{\cos\left[\pi(2j+1)/dz\right]}{\nu^2 + \left[(2j+1)/2d\right]^2},$$

$$E_y = 0, \quad E_z(x,z) \neq 0. \tag{30}$$

Though $E_z \neq 0$, no phase modulation $\psi_1(x)$ through the longitudinal effect arises since $\psi_1(x)$ is proportional to the spatial modulation of the potential difference in the longitudinal effect. In our case, however, the crystal is between two equipotential surfaces and

$$\psi_1(x) \propto \int_{-d/2}^{+d/2} E_z(x,z)\,dz = 0.$$

As in the previous case, the field grating, and hence the refractive index grating, is shifted by a quarter of the spacing for the transverse effect.

The field $E_x(x,z)$ in (30) is represented as a sum of harmonics, i.e., it can be regarded as a sum of fields from a set of gratings with j-dependent spacings. This is not surprising, because the presence of electrodes gives rise to an infinite number of mirror-charge layers, i.e., a periodic structure of the plane charges spaced by 2d. Some characteristic features of the field $E_x(x,z)$ may be analyzed by taking separate terms of (30) as examples. Let us assume j = 0. Then

$$\begin{aligned} E_x(x,z,j=0) &= \frac{\sigma}{\pi\epsilon\epsilon_0 d}\frac{\nu}{\nu^2+(2d)^{-2}}\sin(2\pi\nu x)\cos\left(\frac{2\pi z}{2d}\right) \\ &= E_x(z)\sin(2\pi\nu x). \end{aligned} \qquad (31)$$

For the transverse effect, taking into account (24),

$$\psi_1(j=0) = \frac{\pi}{\lambda}n^3 r_{41}\int_{-d/2}^{+d/2} E_x(z)\,dz = \frac{2\sigma}{\pi\lambda\epsilon\epsilon_0}n^3 r_{41}\frac{\nu}{\nu^2+(2d)^{-2}} \qquad (32)$$

and the diffraction efficiency

$$\eta(j=0) = \left(\frac{\sigma n^3 r_{41}}{\pi\lambda\epsilon\epsilon_0}\right)^2\left(\frac{\nu}{\nu^2+(2d)^{-2}}\right)^2. \qquad (33)$$

As seen from (31) the field grows for low spatial frequencies (ν < l/2d) with the increasing frequency, rather than reducing as in other cases. Accordingly, for the transverse effect the diffraction efficiency tends to zero as $\nu \to 0$ and peaks in the frequency range $\nu \sim$ l/2d. Here the effect arises because of the electrodes present on the crystal plate. However, a similar dependence for the field E_x could be obtained for the grating given by (22) at $\xi = 0$. Equations (26, 31) bear a close resemblance to each other from the point of view of frequency dependence if we assume that $\gamma = (2d)^{-1}$, $\xi = 0$. Equation (31) for E_x can be regarded as resulting from the formation of two charge gratings with K_1 and K_2, where $K_{1x} = K_{2x} = 2\pi\nu$ and $K_{1z} = -K_{2z} = 2\pi\gamma$, and $\gamma = (2d)^{-1}$. Then

$$E_x\left(x,z,j=0\right)=\frac{\rho h}{2\pi\epsilon\epsilon_0 d}\frac{\nu}{\nu^2+\gamma^2}\left(\sin K_1 r+\sin K_2 r\right). \tag{34}$$

Note that, although we are discussing only one term of (30), Eq.(31) proves to give satisfactory results for the limiting values $\nu \ll l/2d$ and $\nu \gg l/2d$ since the term with j = 0 plays a dominating role.

Thus, deposition of electrodes on the crystal plate is qualitatively equivalent - with an accuracy on the order of unity - to the recording of two gratings in the crystal without electrodes, but with the wave vectors having nonzero components not only in the direction of x, but also of z.

7.2.2 Transfer Function

In the framework of linear space-invariant systems theory, spatial light modulators (or simply thin plates of PRCs) can be described by transfer functions or impulse responses. Evidently, because of the anisotropic properties of PRCs, the transfer functions are generally two-dimensional functions of spatial frequencies, and the impulse responses are the two-dimensional functions of coordinates. The impulse response and transfer function of the SLMs establish the relation between the input and output signals.

Let us first discuss what we mean by the input and output signals. For the optically addressed electro-optic SLMs discussed here, the input signal is the magnitude of exposure with recording light

$$W\left(x,y\right)=I\left(x,y\right)\tau_{ex}. \tag{35}$$

The problem of interest in the analysis of coherent optical systems is how the complex amplitude of the readout light changes during its passage through the elements of the optical system. We assume here, with thin holograms, that the complex amplitude of the readout light at the output of a SLM $A_{out}(x, y)$ is related to the incident readout light amplitude A_{in} by $A_{out}(x, y) = A_{in}\, T(x, y)$, where T(x, y) is the complex transmittance of the SLM. Therefore we regard a change of the complex transmittance T(x, y) caused by the recording beam as the output signal. Note once more that T(x, y) relates the complex amplitudes of the readout light, and not its intensities, at the input and output. Since A_{out} and A_{in} can differ by the polarization state, generally speaking, T(x, y) is a tensor that can be written as a 2x2 matrix, because A_{in} and A_{out} are regarded as two-dimensional Maxwell vectors. However, we do not consider the tensorial properties of the transmittance now.

Assume that an image of an infinitesimally small point object with the coordinates x′, y′ is recorded on the SLM. With a corresponding normalization, the input signal is δ(x-x′, y-y′), i.e., a two-dimensional Dirac &-function. Since the SLM, as any other recording medium, has a limited resolution, the changes of its transmittance, i.e., the output signal, will not be localized in an infinitesimally small region.

Let us introduce the function h(x-x′, y-y′) which, with an accuracy to a normalizing factor, is the change of transmittance caused by a δ-function input signal. The function h(x-x′,y-y′) is termed the impulse response of an SLM. If an arbitrary image W(x′y′) is recorded on the modulator, the output signal will be given by the superposition integral

$$T(x, y) = \iint W(x', y')h(x - x', y - y')dx'dy'. \tag{36}$$

Thus, knowledge of the function h(x-x′,y-y′) allows the SLM response to an arbitrary input signal W(x′, y′) to be defined. In practical situations, however, it is common to use the transfer function χ(ν,ξ) which is a Fourier transform of the impulse response at x′ = y′ = 0, i.e.,

$$\chi(\nu,\xi) = \iint h(x, y) exp\left[2\pi i(\nu x + \xi y)\right]dxdy. \tag{37}$$

The transfer function relates the Fourier spectra of the recorded image and the output signal

$$\tilde{T}(\nu,\xi) = \chi(\nu,\xi)\tilde{W}(\nu,\xi), \tag{38}$$

where $\tilde{T}(\nu,\xi)$ and $\tilde{W}(\nu,\xi)$ are the Fourier spectra of T(x, y) and W(x, y), respectively.

Experimental Determination of the Transfer Function

Let us see how the transfer function can be defined using the diffraction technique. The image of a sinusoidal grating is recorded on the modulator

$$W(x, y) = W_0 \left[1 + m\cos 2\pi(\nu x + \xi y)\right]. \tag{39}$$

The grating may be recorded by the holographic method as a pattern of interference between two plane coherent beams or focused on the SLM from a transparency in incoherent light. If we do not take into account the limited sizes of the SLM's active surface, the Fourier spectrum of the recorded grating, (39), according to (38) yields the output signal spectrum density

$$\tilde{T}(\nu,\xi)=W_0\left[\chi(\nu',\xi')\delta(\nu',\xi')+\frac{m}{2}\chi(\nu',\xi')\delta(\nu-\nu',\xi-\xi')+\frac{m}{2}\chi(\nu',\xi')\delta(\nu+\xi',\xi+\xi')\right]. \quad (40)$$

After the grating has been recorded, the diffraction of the readout light transmitted through the modulator is observed. The intensity of the first diffraction order is

$$I_1(\nu,\xi)=\left|\frac{A_{in}\chi(\nu,\xi)W_0 m}{2}\right|^2. \quad (41)$$

Measurements of the diffraction intensity I_1 (ν,ξ) yield the diffraction efficiency η. As (41) reveals, we can find the transfer function modulus by measuring η for gratings with different ν and ξ, but with the same amplitude mW_0

$$\left|\chi(\nu,\xi)\right|=\frac{2}{mW_0}\sqrt{\eta(\nu,\xi)}. \quad (42)$$

Next we must find how $\chi(\nu, \xi)$ is related to the electro-optic and photorefractive properties of the crystal. Generally, it is a fairly involved task if we allow for anisotropic properties of the materials used. The problem was solved for cubic crystals with an arbitrary electric field distribution within the crystal. We take a thin charge grating (Fig. 5) as an example, and assume that the readout light corresponds only to one eigenmode of the electro-optic crystal. Note that this is not a common situation, but a real problem of SLM can be reduced to the superposition of two eigenmodes. Now, using (24, 28) we obtain

$$\chi(\nu\text{ü}\ )=\frac{i\psi_{\text{ü}}}{mW_{\text{ü}}}=\frac{in^3 r\ \sigma}{2\lambda\epsilon\epsilon\ \nu mW} \quad (43)$$

in order to apply a linear approximation, the condition $\psi \ll 1$ should be fullfilled and all terms with $|p| > 1$ should be neglected; The complete solution requires that the relation between σ and W_0 be known, but this relation depends on the particular structure of the device.

As seen from (43), $\chi(\nu, 0) \propto 1/\nu$ for the example under discussion. If the modulator has electrodes on the front and back faces, then, similar to (32), we have $\chi(\nu, 0) \propto \nu$ for $\nu \ll 1/2d$, and $\chi(\nu, 0) \propto 1/\nu$ for $\nu \gg 1/2d$ for the transverse effect. The $\chi(\nu, 0)$ dependence of the longitudinal effect is close to $1/(\nu^2+(2d)^{-2})$.

We have given examples for gratings with K parallel to the axis. If K is arbitrarily oriented in the xy plane, χ is the function $\chi(\nu, \xi)$ of two spatial frequencies. Note that in experiments it is often more convenient to measure χ as a function of the grating wave-vector modulus K and the angle of the vector K with respect to a selected axis. Passing from one variable to the other is not difficult. For instance, the experimental dependence $\chi \propto (K\cos\theta)/[K^2 + (2\pi/2d)^2]$, where θ is the angle between K and the direction of one of the crystal axes, corresponds to $\chi(\nu, \xi) \propto \nu[\nu^2 + \xi^2 + (2d)^{-2}]^{-1}$.

The question frequently arises as to whether it is possible to measure T(x, y) and $\chi(\nu, \xi)$ - i.e., the parameters describing the system that is linear with respect to the readout amplitude - using the methods developed for the incoherent systems that are linear with respect to the readout intensity. It is interesting to consider the relationship between the SLM's coherent transfer function discussed here and the incoherent optical transfer function H(ν, ξ). If we use the incoherent light, we suppose that the optical system is linear with respect to the readout intensity, and the modulator is described by means of the transmittance φ (x, y) such that

$$I_{out}\left(x\ddot{u}y\right) = \phi\left(x \quad y\right) I_{in} \tag{44}$$

where $\varphi(x, y) = |T(x, Y)|^2$, and I_{out}, I_{in} are the readout light intensities at the front and back faces of the device, respectively.

The two conditions, (44) and $A_{out}(x, y) = T(x, y)A_{in}$, require that both T(x, y) and $\varphi(x, y) = |T(x, y)|^2$ be linear functions of the recording light intensity I(x, y). Generally speaking, this is possible if there is a strong bias component of the readout light at the modulator output; that is, either $|\chi(0,0)| \gg \chi(\nu, \xi)$ at ν and $\xi \neq 0$ or there is a beam which is uniform over the modulator aperture and whose intensity is well above that of the diffracted beam. Then we can introduce the incoherent transfer function H(ν, ξ), which is related to the coherent transfer function $\chi(\nu, \xi)$ by

$$\mathcal{H}\left(\nu, \xi\right) = \iint \chi^*\left(\nu + \nu', \xi + \xi'\right) \chi\left(\nu', \xi'\right) d\nu' d\xi'. \tag{45}$$

Since H(ν, ξ) is the autocorrelation function of $\chi(\nu, \xi)$, it peaks at $\nu = \xi = 0$. Therefore normalizing it to H(0, 0) is appropriate. The modulus of the relation

$$\left|\frac{\mathcal{H}\left(\nu, \xi\right)}{\mathcal{H}\left(0, 0\right)}\right| = MTF \tag{46}$$

is called the modulation transfer function. The MTF is typically measured as the contrast of the recorded grating versus frequency. This discussion clearly shows that the MTF can be unambiguously defined for a system that is linear with respect to both the amplitude and intensity of the readout light if $\chi(\nu, \xi)$ is known, but not vice versa. Also, $|\chi(\nu, \xi)|$ is readily found from the MTF measurements only at low contrasts and for $\chi(0, 0) \gg \chi(\nu, \xi)$.

We note in conclusion that, when used with coherent light, the electrooptic modulators are operated in such a mode that A_{out} and A_{in} have different polarization states. Therefore, in practical situations an analyzer of the polarization is commonly placed behind the SLM to suppress the undiffracted part of the light to reduce the noise.

7.2.3 Dynamic Range and Information Capacity

The dynamic range of photorefractive media is still not understood fully despite the fact that the relatively low values of the dynamic range limit applicability of PRCs. Formally, the dynamic range D can be characterized by the ratio

$$D = 10lg\frac{I_{max}}{I_{min}}, \quad (47)$$

where I_{min} is the minimum level of the usable optical signal that can be detected during readout of the information recorded in the PRC, and I_{max} is the maximum level of the readout signal within the linear region of the input-output signal relation (by a certain specified criterion). Generally, I_{min} is determined by the noise level due to scattered light and is related to the non-homogeneities, i.e., defects of the photorefractive medium. Rather often the minimum level is determined by other optical elements incorporated in the optical scheme or by the sensitivity of the photodetector.

On the other hand, I_{max} is limited either by the nonlinearities of the recording mechanism or the number of photoactive centers in the crystal. For instance, as will be shown for SLMs, the field induced in the crystal during recording can strongly affect the recording process itself and give rise to higher-order spatial harmonics and nonlinear distortions. A limited concentration of donors and traps is essential for recording holograms at high spatial frequencies. A sufficiently accurate theoretical evaluation of the dynamic range cannot be made since the noise level can be measured only in experiments. However, situations are frequently encountered in which even the available experimental evidence does not provide an estimate of the dynamic range for an arbitrary form of the input signal.

The fact is that the experimental data depend on whether the dynamic range is measured in the image or Fourier plane and on how the spatial frequency bandwidth is controlled during measurements. For instance, if measurements are carried out in the Fourier plane and the sizes of the photodetector correspond to the diffraction-limited spot, then, if a sinusoidal grating is recorded at the input, the signal-to-noise ratio in the Fourier plane will approximately be N times as high as that in the input plane. This is due to the focusing of the light energy into a small spot. Here $N \simeq (2L\Delta\nu)^2$, where L is the size of the input aperture and $\Delta\nu$ is the transmitted frequency bandwidth. We assume that we have white noise and the signal-to-noise ratio in the input plane is measured at the spot of $(1/\Delta\nu)^2$ in size. If a noise-like signal is recorded and the conditions are those given above, the dynamic ranges in the image plane and the Fourier plane must be nearly the same.

Thus, the dynamic range of the measuring techniques based upon diffraction is a function of the input signal, the input aperture, and the bandwidth. Recalculating the data obtained for a single grating to apply to the case of a complex signal is possible in general. However, since certain measurement conditions are not known, it leads to fairly large errors in the estimate of D, and the definition of the information capacity of holograms and SLMs.

The information capacity is an extremely important parameter for an evaluation of the information properties of an individual optical element, and a comparison of the information processing system as a whole. In evaluating the information capacity we can encounter the same difficulties as those arising in the determination of the dynamic range; i.e., the result can differ noticeably depending on where the measurements are carried out (the image plane or the Fourier plane) and on the type of signal for which the analysis is performed. Moreover, confusion sometimes arises in experimental studies since the measured values are not always consistent with those required for calculation.

Let us estimate the information capacity of an SLM using the measurements in the Fourier plane as an example. From familiar theory, the information capacity of a signal recorded on a square-shaped modulator with the linear size L is given by

$$C = \frac{L^2}{2} \iint \log_2 \left(1 + \frac{P_s(\nu,\xi)}{P_n(\nu,\xi)} \right) d\nu d\xi, \tag{48}$$

where $P_s(\nu,\xi)$ and $P_n(\nu,\xi)$ are the spectral densities of the signal and noise power, respectively. The factor 1/2 takes into account the loss of information on phase when the intensity of the Fourier spectrum is measured in the Fourier plane. Note that the value of C depends on the type of the signal $P_s(\nu,\xi)$. Therefore, care is needed in inferring the highest possible information capacity of the device from the experimental data obtained for a specific type of the signal. For instance, for recording a single cosinusoidal grating $\cos[2\pi(\nu_0 x+\xi_0 y)]$ and measuring in the Fourier plane,

$$C \simeq \log_2 \left(1 + \frac{I_s(\nu_0,\xi_0)}{I_n(\nu_0,\xi_0)} \right), \tag{49}$$

where $I_s(\nu_0,\xi_0)$ is the magnitude of the signal, and $I_n(\nu_0,\xi_0)$ is the average noise level in the neighborhood of point ν_0,ξ_0. Note that the information capacity is not high and is determined essentially by the dynamic range of the modulator.

In another situation, if a signal with a rich spectrum, for instance, a noise-like signal, is recorded

$$C \simeq \frac{N}{2} \cdot \log_2 \left(1 + I_{sn}/I_n \right), \tag{50}$$

where I_{sn} and I_n are the mean values of the signal and noise in the Fourier plane, respectively. For simplicity, (50) assumes a constant transfer function of the modulator within the bandwidth and that the noise is white.

Equation (50) can be used to estimate the maximum information capacity of the modulator; i.e., we can state that

$$C_{\max} \simeq N \log_2 \left(1 + I_{sn}/I_n \right). \tag{51}$$

However, experimental studies of the noiselike signal are rather complicated. The ratio I_{sn}/I_n in (50) is sometimes replaced by $I_s(\nu_0,\xi_0)/I_n(\nu_0,\xi_0)$ obtained from measurements of a sinusoidal signal, which gives the overestimated values of C_{max}. Since we can take $I_s(\nu_0,\xi_0) \simeq NI_{sn}$ for the sinusoidal and noise-like signals, the error can be very large and the estimate of C_{max} will be incorrect.

In practice, the inherent noises of the modulator or the detecting device can be so high that it is impossible to detect the noise-like signal with the spectrum equal to Δν, i.e., $I_{sn}/I_n < 1$, in spite of the fact that on recording a single sinusoidal signal,

$$I_{sn}\left(\nu_0,\xi_0\right)/I_n\left(\nu_0,\xi_0\right)>1.$$

Here detection of the noise-like signal - but with a spectrum narrower than Δν - is possible, and then the maximum information capacity can be estimated as

$$C_{\max}\simeq\left(2L\Delta\nu_{eff}\right)^2\simeq\frac{I_s^{\max}\left(\nu_0,\xi_0\right)}{I_n\left(\nu_0,\xi_0\right)},\tag{52}$$

where $\Delta\nu_{eff}$ is the maximum frequency spectrum ($\Delta\nu_{eff} < \Delta\nu$) in the initial noise-like signal that can be detected by the device, and $I_s^{\max}\left(\nu_0,\xi_0\right)$ is the highest possible magnitude of the signal on recording a single sinusoidal grating. Thus, (52) shows that the dynamic range of the modulator is a deciding factor in the latter case.

Limitations imposed by the dynamic range are essential for the volume holograms as well. Though formally $C \sim \Delta\nu^3L^2d$ for a volume hologram, the noises of the crystal itself and the photodetecting system that limit the dynamic range prevent realization of the theoretically possible information capacity. In practice, $C_{max} \sim 10^5$ -10^7 bit/cm^2 for thin holograms and SLMs with an area of 1 cm^2, while for volume holograms it is one to two orders of magnitude higher.

7.2.4 Sensitivity

The sensitivity characterizes the light energy needed to record information in a photosensitive material. Different researchers have defined sensitivity in different ways.

A parameter of great importance is the holographic sensitivity

$$S_h=\frac{\sqrt{\eta}}{Wm},\tag{53}$$

where W is the total light energy (i.e., the sum of the energies of the reference and object beams) required for recording a single grating on an area of 1 cm^2, and m is the visibility (contrast) of the interference fringes. Holographic sensitivity is measured in cm^2 /J. The diffraction efficiency of volume holograms depends on the sample thickness and the Bragg angle, i.e., the conditions of measurements. Therefore, the sensitivity of photosensitive material in volume holograms is sometimes defined by S_h per unit length. Instead of S_h, the reciprocal value S_h^{-1} is often regarded as the holographic sensitivity.

The holographic media are as a rule insufficiently linear with respect to the recording light intensity, and S_h thus depends on W. The definition (53) is often used, but taken for a definite value of the diffraction efficiency, typically $\eta = 1\%$. Here sensitivity is the energy density W needed to achieve $\eta = 0.01$ at m = 1, i.e.,

$$1/S = W(1\%). \tag{54}$$

The sensitivity for linear media (in terms of S_h as a function of $\sqrt{\eta}$) is $1/S = 0.1/S_h$.

In the framework of information theory, the given definitions suffer from the limitation that they do not account for the reconstructed frequency bandwidth and the dynamic range of the medium. For comparing the estimates for the sensitivities of different media in this sense, it is more convenient to express sensitivity as the energy required to record a pixel or a bit of information. However, because of the difficulties in precise evaluation of the information capacity from the experimental data, this definition is seldom used. The sensitivity for SLMs is sometimes defined as the energy required to achieve phase modulation of $\psi_1 = \pi/2$.

Once the expressions for the diffraction efficiency on information recording are known, the equations describing the fundamental dependences of sensitivity are easily derived. Without examining specific cases, we give here the typical expressions for the sensitivity of photorefractive media in the generalized form for optimum charge-transport lengths and in the linear approximation (when $\sqrt{\eta} \propto W$)

$$1/S_h = \frac{\hbar\omega\epsilon\epsilon_0}{n^3 r_{ij}} \frac{1}{\alpha\beta} \frac{1}{m} F(\nu, d, \theta_B). \tag{55}$$

Equation (55) shows that sensitivity becomes lower with the increasing energy of the photon quantum ħw and increasing dielectric permeability, and becomes higher with increasing r_{ij}, n, absorption coefficient α, quantum efficiency β, and modulation depth m. The dependence on the absorption coefficient α is obtained under the assumption that $\alpha d < 1$, where d is the sample thickness. Function $F(\nu, d, \theta_B)$ describes the dependence on spatial frequencies. This dependence takes into account the specific features of the recording and readout geometries used, the role of sample thickness and the type of electro-optic effect.

Note the following curious fact. Equation (55) includes the ratio $\epsilon\epsilon_0/r_{ij}$. But it is known from the data on the electro-optic properties of ox1de crystals that this relation exhibits a weak dependence on both the type of crystal and the temperature. Thus in reality the sensitivity of photorefractive media is slightly affected by their electro-optic properties.

In practice, the observed sensitivities 1/S of the presently known crystals lie in the range 10 to 10^{-3} J/cm^2 at high spatial frequencies ν ~ 300 to 1000 line pairs/mm. The sensitivity of SLMs is determined as in the case of thin holograms. For ν ~ 5 to 10 line pairs/mm, the SLM sensitivity 1/S reaches 10^{-6} J/cm^2.

Estimates of sensitivity per pixel or bit of information yield more universal values that in favorable situations can amount to 10^{-10} to 10^{-11} J, irrespective of whether a volume or a thin grating is recorded.

7.2.5 Speed of Operation and Storage Time

The speed of operation of photosensitive elements fabricated from PRCs is determined by the rates of information recording and erasure. Two most typical cases can be singled out for holographic recording.

First Typical Case

The drift and diffusion lengths are much shorter than the grating spacing. The characteristic time for the charge grating formation is then governed by the time required for the photoexcited carriers to redistribute, i.e., Maxwell's relaxation time τ_M. The characteristic relaxation rate $1/\tau_M$ is determined by the dielectric permeability and conductivity of the crystal σ. With acceptable accuracy,

$$\frac{1}{\tau_M} = \frac{\sigma}{\epsilon\epsilon_0} = \frac{\sigma_D}{\epsilon\epsilon_0} + \frac{\sigma_I}{\epsilon\epsilon_0}, \qquad (56)$$

where σ_D is the conductivity at dark, and $\sigma_I = e\beta\alpha I_0\mu\tau/\hbar w$ is photoconductivity. The process of grating formation is here equivalent to recharging a capacitor with the time constant τ_M.

It is apparent from (56) that τ_M depends on the average light intensity incident on the crystal. Therefore, during recording, $\sigma_I \gg \sigma_D$ and $1/\tau_M \propto I_0$. After the recording light is switched off, erasure at a rate $1/\tau_M \propto \sigma_D$

occurs. The erase rate determines the information storage time of the crystal. To achieve a more rapid erasure, the crystal can be uniformly irradiated with the erase light I_{er}. The erase rate obeys the same relation (56), with I_0 being replaced by I_{er}. The process of erasure is equivalent to capacitor discharging. Note that it is convenient to short-circuit electrodes during accelerated erasure if the drift recording mechanism is used.

The storage time can sometimes be increased through special fixing techniques. In practice, storage times of PRCs at room temperature lie in a wide range, from microseconds to days.

Second Typical Case

The drift and diffusion lengths are well above the grating spacing. Here the charge-grating formation rate is primarily determined by the electron photoexcitation rate, since their subsequent redistribution scarcely affects the charge grating produced by the positively ionized donors.

The mechanisms governing the recording and erase rates during information recording on SLMs are basically the same. However, the most important parameter for SLMs is the allowable record-read-erase cycle rate. It can prove to be much less than $1/\tau_M$, since multiple recording and erasure processes give rise to heat release, causing overheating of the modulator. In practice, the cycle rate of such modulators as PROM and PRIZ is several tens of Hz, while τ_M can reach 10^{-3} to 10^{-6} s.

7.2.6 Resolution

Resolution is an important parameter of photosensitive materials, but, unfortunately, defined in different ways. In conventional optics, resolution is often determined as the size of the minimum element that can be detected by the photosensitive material on recording a test object, for instance, a resolution chart. This method is useful for non-coherent optical systems, but it is less attractive for coherent ones, since here more suitable diffraction methods can be used.

In holography (and, generally, coherent optical systems) resolution is typically defined as the bandwidth of spatial frequencies $\Delta\nu$ within which the transfer function $\chi(\nu)$ or diffraction efficiency $\eta(\nu)$ decreases by a certain factor. Although this is common practice, the definition is confusing, because, first, different researches use different criteria (a decrease of $\eta(\nu)$ by a factor of 2, 3, 10 etc.) and, second, the dynamic range or noise of the material itself are not taken into account. Therefore the resolution obtained cannot be used for estimation of the information capacity of the material. Diffraction efficiencies of individual gratings as functions of ν can be measured in sequence in a wide range of spatial frequencies and, formally, the bandwidth $\Delta\nu$ can be fairly large. However, if a large number of gratings is recorded simultaneously (for a complicated image}, not all these gratings can be simultaneously reconstructed because of a definite level of noise and a limited dynamic range of the photosensitive material. Thus, in practice, the effective bandwidth $\Delta\nu_{eff}$ can be much smaller than $\Delta\nu$. For instance, for thin holograms or SLMs using photorefractive materials, individual gratings with spatial frequencies up to hundreds of lines per millimeter can be recorded. However, $\Delta\nu_{eff}$ that provides a simultaneous reconstruction of a set of gratings with an appropriate signal-to-noise ratio is 15-20 lines/mm. The maximum spatial frequency of an individual grating and bandwidth $\Delta\nu_{eff}$ for volume holograms are several times higher than for thin holograms, because volume holograms have higher diffraction efficiencies. In spite of the fact that $\Delta\nu_{eff}$ is the most appropriate characteristic for information processing systems, literature rarely provides data on the effective bandwidth.

SUMMARY

- Photorefraction is an effect of refractive index variation in the material exposed to the light. This property makes it possible to create spatial distribution of refractive index in the crystal and can be used in wide variety of devices including spatial light modulators, optical memory, holography, phase conjugation etc.
- The photorefractive effect is a nonlinear optical effect seen in certain crystals and other materials that respond to light by altering their refractive index. The effect can be used to store temporary, erasable holograms and is useful for holographic data storage. It can also be used to create a phase-conjugate mirror or an optical spatial soliton.
- The field grating is shifted by $\pi/2$ with respect to the interference pattern (cos($K_x x$) has been transformed into sin($K_x x$)] and the grating amplitude is proportional to the spatial frequency K_x. Moreover, comparison of (6 and 3) also reveals that the buildup rate of the field grating grows with increasing L_d and decreasing ϵ, with the recording light intensity being the same.
- An electro-optic effect is a change in the optical properties of a material in response to an electric field that varies slowly compared with the frequency of light. The term encompasses a number of distinct phenomena, which can be subdivided into changes in absorption and changes in refractive index.
- A non-uniform charge - and thus electric field - distribution induced in a photorefractive crystal cause variation of the index of refraction through the electro-optic effect.
- The electro-optic effect consists of a change of the refractive index of the crystal and a change of the orientation of its optical axes on application of an external electric field.
- The simplest relations are obtainable for an originally isotropic (cubic) crystal. With n_0 applied field, $n_0 = n_e = n$ and $\varphi_{oe} = 0$. If the field E_ℓ is applied along one of the cubic crystallographic axes-for instance, the [001] direction - the crystal becomes a biaxial one with $n_z = n$, $n_{x'} = n+\Delta n$, $n_{y'} = n-\Delta n$, where $\Delta n = ½n^3\ r_{41}\ E_\ell d$. We assume here that the z axis || [001], and the two other axes x′ and y′ are at ±45° to the (100] axis.

- The mechanism of information recording determines such characteristics of the photorefractive crystals as diffraction efficiency, information capacity, spatial-frequency bandwidth, sensitivity, speed of operation, dynamic range, and storage time.
- Evaluating the diffraction efficiency involves calculation of the amplitude of refractive index variation Δn. Let us restrict our analysis to one eigenmode of the light wave in the crystal. Then Δn denotes the refractive index deviation for a given mode (for instance, ordinary or extraordinary beam). Because Δn is unambiguously related to the grating field amplitude through the electro-optic coefficient, the field in the crystal must be determined for a specified charge distribution.
- The dynamic range of photorefractive media is still not understood fully despite the fact that the relatively low values of the dynamic range limit applicability of PRCs.

REFERENCES

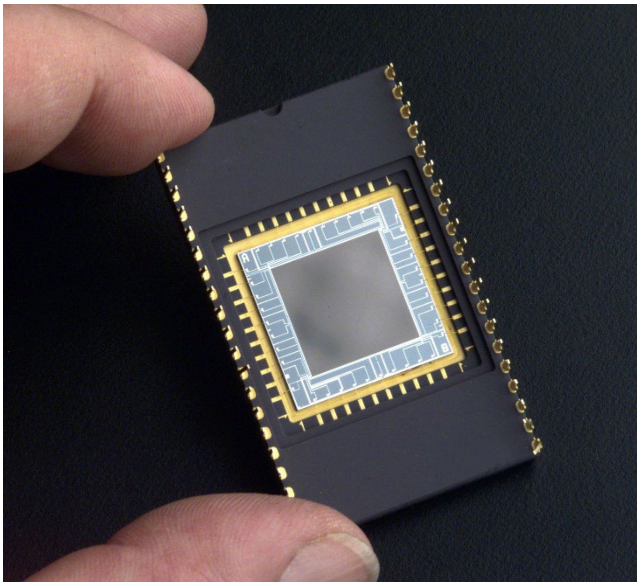
1. Castles, F. (2015-12-03). "Linear electro-optic effects due to high-order spatial dispersion". Physical Review A. American Physical Society (APS). 92 (6): 063804. arXiv:1503.04103. doi:10.1103/physreva.92.063804. ISSN 1050-2947.
2. Consoli, F.; De Angelis, R.; Duvillaret, L.; Andreoli, P. L.; Cipriani, M.; Cristofari, G.; Di Giorgio, G.; Ingenito, F.; Verona, C. (15 June 2016). "Time-resolved absolute measurements by electro-optic effect of giant electromagnetic pulses due to laser-plasma interaction in nanosecond regime". Scientific Reports. 6 (1): 27889. Bibcode:2016NatSR...627889C. doi:10.1038/srep27889. PMC 4908660. PMID 27301704.
3. J. Frejlich (2007). Photorefractive materials: fundamental concepts, holographic recording and materials characterization. ISBN 978-0-471-74866-3.
4. Peter Günter, Jean-Pierre Huignard, ed. (2007). Photorefractive materials and their applications. ISBN 978-0-387-34443-0.
5. Robinson, T. S.; Consoli, F.; Giltrap, S.; Eardley, S. J.; Hicks, G. S.; Ditter, E. J.; Ettlinger, O.; Stuart, N. H.; Notley, M.; De Angelis, R.; Najmudin, Z.; Smith, R. A. (20 April 2017). "Low-noise time-resolved optical sensing of electromagnetic pulses from petawatt laser-matter interactions". Scientific Reports. 7 (1): 983. Bibcode:2017NatSR...7..983R. doi:10.1038/s41598-017-01063-1. PMC 5430545. PMID 28428549.

CHAPTER 8

CHARGE COUPLED DEVICES (CCD)

INTRODUCTION

A charge-coupled device (CCD) is an integrated circuit containing an array of linked, or coupled, capacitors. Under the control of an external circuit, each capacitor can transfer its electric charge to a neighboring capacitor. CCD sensors are a major technology used in digital imaging.



In a CCD image sensor, pixels are represented by p-doped metal–oxide–semiconductor (MOS) capacitors. These MOS capacitors, the basic building blocks of a CCD, are biased above the threshold for inversion when image acquisition begins, allowing the conversion of incoming photons into electron charges at the semiconductor-oxide interface; the CCD is then used to read out these charges. Although CCDs are not the only technology to allow for light detection, CCD image sensors are widely used in professional, medical, and scientific applications where high-quality image data are required. In applications with less exacting quality demands, such as consumer and professional digital cameras, active pixel sensors, also known as CMOS sensors (complementary MOS sensors), are generally used. However, the large quality advantage CCDs enjoyed early on has narrowed over time and since the late 2010s CMOS sensors are the dominant technology, having largely if not completely replaced CCD image sensors.

8.1 CONCEPT OF CHARGE COUPLED DEVICES

A charge coupled device (CCD) is a highly sensitive photon detector. The CCD is divided up into a large number of light-sensitive small areas (known as pixels) which can be used to build up an image of the scene of interest. A photon of light which falls within the area defined by one of the pixels will be converted into one (or more) electrons and the number of electrons collected will be directly proportional to the intensity of the scene at each pixel. When the CCD is clocked out, the number of electrons in each pixel are measured and the scene can be reconstructed.

Digital camera systems, incorporating a variety of charge-coupled device (CCD) detector configurations, are by far the most common image capture technology employed in modern optical microscopy. Until recently, specialized conventional film cameras were generally used to record images observed in the microscope. This traditional method, relying on the photon-sensitivity of silver-based photographic film, involves temporary storage of a latent image in the form of photochemical reaction sites in the exposed film, which only becomes visible in the film emulsion layers after chemical processing (development).

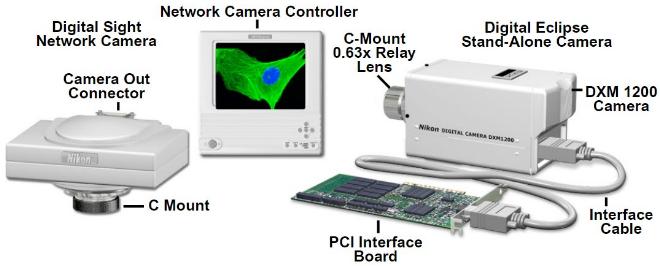


Figure 1. Digital CCD camera systems for optical microscopy.

Digital cameras replace the sensitized film with a CCD photon detector, a thin silicon wafer divided into a geometrically regular array of thousands or millions of light-sensitive regions that capture and store image information in the form of localized electrical charge that varies with incident light intensity. The variable electronic signal associated with each picture element (pixel) of the detector is read out very rapidly as an intensity value for the corresponding image location, and following digitization of the values, the image can be reconstructed and displayed on a computer monitor virtually instantaneously.

Several digital camera systems designed specifically for optical microscopy are illustrated in Figure 1. The Nikon Digital Eclipse DXM1200 provides high quality photo-realistic digital images at resolutions ranging up to 12 million pixels with low noise, superb color rendition, and high sensitivity. The camera is controlled by software that allows the microscopist a great deal of latitude in collecting, organizing, and correcting digital images. Live color monitoring on the supporting computer screen at 12 frames per second enables easy focusing of images, which can be saved with a choice of three formats: JPG, TIF, and BMP for greater versatility.

The DS-5M-L1 Digital Sight camera system is innovative digital imaging system for microscopy that emphasizes the ease and efficiency of an all-in-one concept, incorporating a built-in LCD monitor in a stand-alone control unit. The system optimizes the capture of high-resolution images up to 5 megapixels through straightforward menus and pre-programmed imaging modes for different observation methods. The stand-alone design offers the advantage of independent operation including image storage to a CompactFlash Card housed in the control/monitor unit, but has the versatility of full network capabilities if desired. Connection is possible to PCs through a USB interface, and to local area networks or the Internet via Ethernet port. Web browser support is available for live image viewing and remote camera control, and the camera control unit supports HTTP, Telnet, FTP server/client, and is DHCP compatible. The camera systems illustrated in Figure 1 represent the advanced technology currently available for digital imaging with the optical microscope.

Perhaps the single most significant advantage of digital image capture in optical microscopy, as exemplified by CCD camera systems, is the possibility for the microscopist to immediately determine whether a desired image has been successfully recorded. This capability is especially valuable considering the experimental complexities of many imaging situations and the transient nature of processes that are commonly investigated. Although the charge-coupled device detector functions in an equivalent role to that of film, it has a number of superior attributes for imaging in many applications. Scientific-grade CCD cameras exhibit extraordinary dynamic range, spatial resolution, spectral bandwidth, and acquisition speed. Considering the high light sensitivity and light collection efficiency of some CCD systems, a film speed rating of approximately ISO 100,000 would be required to produce images of comparable signal-to-noise ratio (SNR). The spatial resolution of current CCDs is similar to that of film, while their resolution of light intensity is one or two orders of magnitude better than that achieved by film or video cameras. Traditional photographic films exhibit no sensitivity at wavelengths exceeding 650 nanometers in contrast to high-performance CCD sensors, which often have significant quantum efficiency into the near infrared spectral region. The linear response of CCD cameras over a wide range of light intensities contributes to the superior performance, and gives such systems quantitative capabilities as imaging spectrophotometers.

A CCD imager consists of a large number of light-sensing elements arranged in a two-dimensional array on a thin silicon substrate. The semiconductor properties of silicon allow the CCD chip to trap and hold photon-induced charge carriers under appropriate electrical bias conditions. Individual picture elements, or pixels, are defined in the silicon matrix by an orthogonal grid of narrow transparent current-carrying electrode strips, or gates, deposited on the chip. The fundamental light-sensing unit of the CCD is a metal oxide semiconductor (MOS) capacitor operated as a photodiode and storage device. A single MOS device of this type is illustrated in Figure 2, with reverse bias operation causing negatively charged electrons to migrate to an area underneath the positively charged gate electrode. Electrons liberated by photon interaction are stored in the depletion region up to the full well reservoir capacity. When multiple detector structures are assembled into a complete CCD, individual sensing elements in the array are segregated in one dimension by voltages applied to the surface electrodes and are electrically isolated from their neighbors in the other direction by insulating barriers, or channel stops, within the silicon substrate.

The light-sensing photodiode elements of the CCD respond to incident photons by absorbing much of their energy, resulting in liberation of electrons, and the formation of corresponding electron-deficient sites (holes) within the silicon crystal lattice. One electron-hole pair is generated from each absorbed photon, and the resulting charge that accumulates in each pixel is linearly proportional to the number of incident photons. External voltages applied to each pixel's electrodes control the storage and movement of charges accumulated during a specified time interval. Initially, each pixel in the sensor array functions as a potential well to store the charge during collection, and although either negatively charged electrons or positively charged holes can be accumulated (depending on the CCD design), the charge entities generated by incident light are usually referred to as photoelectrons. This discussion considers electrons to be the charge carriers. These photoelectrons can be accumulated and stored for long periods of time before being read from the chip by the camera electronics as one stage of the imaging process.

Image generation with a CCD camera can be divided into four primary stages or functions: charge generation through photon interaction with the device's photosensitive region, collection and storage of the liberated charge, charge transfer, and charge measurement. During the first stage, electrons and holes are generated in response to incident photons in the depletion region of the MOS capacitor structure, and liberated electrons migrate into a potential well-formed beneath an adjacent positively-biased gate electrode. The system of aluminum or polysilicon surface gate electrodes overlie, but are separated from, charge carrying channels that are buried within a layer of insulating silicon dioxide placed between the gate structure and the silicon substrate. Utilization of polysilicon as an electrode material provides transparency to incident wavelengths longer than approximately 400 nanometers and increases the proportion of surface area of the device that is available for light collection. Electrons generated in the depletion region are initially collected into electrically positive potential wells associated with each pixel. During readout, the collected charge is subsequently shifted along the transfer channels under the influence of voltages applied to the gate structure. Figure 3 illustrates the electrode structure defining an individual CCD sense element.

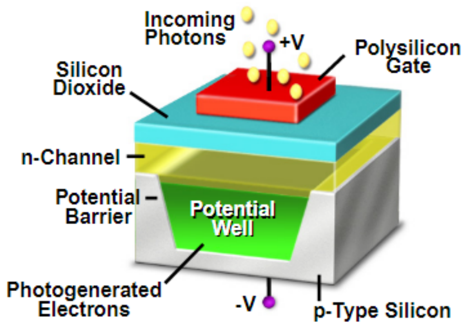


Figure 2. Metal oxide semiconductor (MOS) capacitor.

In general, the stored charge is linearly proportional to the light flux incident on a sensor pixel up to the capacity of the well; consequently this full-well capacity (FWC) determines the maximum signal that can be sensed in the pixel, and is a primary factor affecting the CCD's dynamic range. The charge capacity of a CCD potential well is largely a function of the physical size of the individual pixel. Since first introduced commercially, CCDs have typically been configured with square pixels assembled into rectangular area arrays, with an aspect ratio of 4:3 being most common. Figure 4 presents typical dimensions of several of the most common sensor formats in current use, with their size designations in inches according to a historical convention that relates CCD sizes to vidicon tube diameters.

8.1.1 CCD Formats

The rectangular geometry and common dimensions of CCDs result from their early competition with vidicon tube cameras, which required the solid-state sensors to produce an electronic signal output that conformed to the prevailing video standards at the time. Note that the "inch" designations do not correspond directly to any of the CCD dimensions, but represent the size of the rectangular area scanned in the corresponding round vidicon tube. A designated "1-inch" CCD has a diagonal of 16 millimeters and sensor dimensions of 9.6 x 12.8 millimeters, derived from the scanned area of a 1-inch vidicon tube with a 25.4-millimeter outside diameter and an input window approximately 18 millimeters in diameter. Unfortunately, this confusing nomenclature has persisted, often used in reference to CCD "type" rather than size, and even includes sensors classified by a combination of fractional and decimal terms, such as the widely used 1/1.8-inch CCD that is intermediate in size between 1/2-inch and 2/3-inch devices.

Although consumer cameras continue to primarily employ rectangular sensors built to one of the "standardized" size formats, it is becoming increasingly common for scientific-grade cameras to incorporate square sensor arrays, which better match the circular image field projected in the microscope. A large range of sensor array sizes are produced, and individual pixel dimensions vary widely in designs optimized for different performance parameters. CCDs in the common 2/3-inch format typically have arrays of 768 x 480 or more diodes and dimensions of 8.8 x 6.6 millimeters (11-millimeter diagonal). The maximum dimension represented by the diagonal of many sensor arrays is considerably smaller than the typical microscope field of view, and results in a highly magnified view of only a portion of the full field. The increased magnification can be beneficial in some applications, but if the reduced field of view is an impediment to imaging, demagnifying intermediate optical components are required. The alternative is use of a larger CCD that better matches the image field diameter, which ranges from 18 to 26 millimeters in typical microscope configurations.

An approximation of CCD potential-well storage capacity may be obtained by multiplying the diode (pixel) area by 1000. A number of consumer-grade 2/3-inch CCDs, with pixel sizes ranging from 7 to 13 micrometers in size, are capable of storing from 50,000 to 100,000 electrons. Using this approximation strategy, a diode with 10 x 10 micrometer dimensions will have a full-well capacity of approximately 100,000 electrons. For a given CCD size, the design choice regarding total number of pixels in the array, and consequently their dimensions, requires a compromise between spatial resolution and pixel charge capacity. A trend in current consumer devices toward maximizing pixel count and resolution has resulted in very small diode sizes, with some of the newer 2/3-inch sensors utilizing pixels less than 3 micrometers in size.

CCDs designed for scientific imaging have traditionally employed larger photodiodes than those intended for consumer (especially video-rate) and industrial applications. Because full-well capacity and dynamic range are direct functions of diode size, scientific-grade CCDs used in slow-scan imaging applications have typically employed diodes as large as 25 x 25 micrometers in order to maximize dynamic range, sensitivity, and signal-to-noise ratio. Many current high-performance scientific-grade cameras incorporate design improvements that have enabled use of large arrays having smaller pixels, which are capable of maintaining the optical resolution of the microscope at high frame rates. Large arrays of several million pixels in these improved designs can provide high-resolution images of the entire field of view, and by utilizing pixel binning and variable readout rate, deliver the higher sensitivity of larger pixels when necessary.

8.1.2 Readout of CCD Array Photoelectrons

Before stored charge from each sense element in a CCD can be measured to determine photon flux on that pixel, the charge must first be transferred to a readout node while maintaining the integrity of the charge packet. A fast and efficient charge-transfer process, as well as a rapid readout mechanism, are crucial to the function of CCDs as imaging devices. When a large number of MOS capacitors are placed close together to form a sensor array, charge is moved across the device by manipulating voltages on the capacitor gates in a pattern that causes charge to spill from one capacitor to the next, or from one row of capacitors to the next. The translation of charge within the silicon is effectively coupled to clocked voltage patterns applied to the overlying electrode structure, the basis of the term "charge-coupled" device. The CCD was initially conceived as a memory array, and intended to function as an electronic version of the magnetic bubble device. The charge transfer process scheme satisfies the critical requirement for memory devices of establishing a physical quantity that represents an information bit, and maintaining its integrity until readout. In a CCD used for imaging, an information bit is represented by a packet of charges derived from photon interaction. Because the CCD is a serial device, the charge packets are read out one at a time.

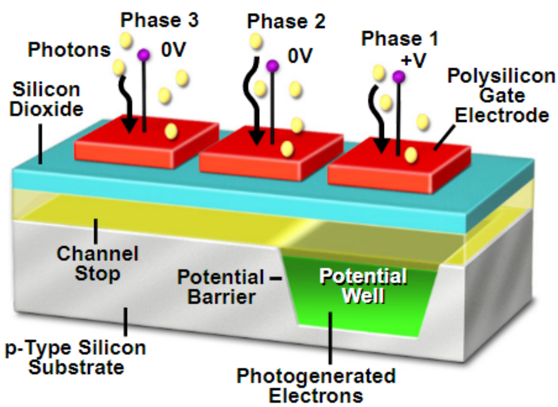


Figure 3. CCD Sense element (pixel) structure.

The stored charge accumulated within each CCD photodiode during a specified time interval, referred to as the integration time or exposure time, must be measured to determine the photon flux on that diode. Quantification of stored charge is accomplished by a combination of parallel and serial transfers that deliver each sensor element's charge packet, in sequence, to a single measuring node. The electrode network, or gate structure, built onto the CCD in a layer adjoining the sensor elements, constitutes the shift register for charge transfer. The basic charge transfer concept that enables serial readout from a two-dimensional diode array initially requires the entire array of individual charge packets from the imager surface, constituting the parallel register, to be simultaneously transferred by a single-row incremental shift. The charge-coupled shift of the entire parallel register moves the row of pixel charges nearest the register edge into a specialized single row of pixels along one edge of the chip referred to as the serial register. It is from this row that the charge packets are moved in sequence to an on-chip amplifier for measurement. After the serial register is emptied, it is refilled by another row-shift of the parallel register, and the cycle of parallel and serial shifts is repeated until the entire parallel register is emptied. Some CCD manufacturers utilize the terms vertical and horizontal in referring to the parallel and serial registers, respectively, although the latter terms are more readily associated with the function accomplished by each.

A widely used analogy to aid in visualizing the concept of serial readout of a CCD is the bucket brigade for rainfall measurement, in which rain intensity falling on an array of buckets may vary from place to place in similarity to incident photons on an imaging sensor. The parallel register is represented by an array of buckets, which have collected various amounts of signal (water) during an integration period. The buckets are transported on a conveyor belt in stepwise fashion toward a row of empty buckets that represent the serial register, and which move on a second conveyor oriented perpendicularly to the first. In Figure 5(b), an entire row of buckets is being shifted in parallel into the reservoirs of the serial register. The serial shift and readout operations are illustrated in Figure 5(c), which depicts the accumulated rainwater in each bucket being transferred sequentially into a calibrated measuring container, analogous to the CCD output amplifier. When the contents of all containers on the serial conveyor have been measured in sequence, another parallel shift transfers contents of the next row of collecting buckets into the serial register containers, and the process repeats until the contents of every bucket (pixel) have been measured.

There are many designs in which MOS capacitors can be configured, and their gate voltages driven, to form a CCD imaging array. Gate electrodes are arranged in strips covering the entire imaging surface of the CCD face. The simplest and most common charge transfer configuration is the three-phase CCD design, in which each photodiode (pixel) is divided into thirds with three parallel potential wells defined by gate electrodes. In this design, every third gate is connected to the same clock driver circuit. The basic sense element in the CCD, corresponding to one pixel, consists of three gates connected to three separate clock drivers, termed phase-1, phase-2, and phase-3 clocks. Each sequence of three parallel gates makes up a single pixel's register, and the thousands of pixels covering the CCD's imaging surface constitute the device's parallel register. Once trapped in a potential well, electrons are moved across each pixel in a three-step process that shifts the charge packet from one pixel row to the next. A sequence of voltage changes applied to alternate electrodes of the parallel (vertical) gate structure move the potential wells and the trapped electrons under control of a parallel shift register clock. The general clocking scheme employed in three-phase transfer begins with a charge integration step, in which two of the three parallel phases per pixel are set to a high bias value, producing a high-field region relative to the third gate, which is held at low or zero potential. For example, phases 1 and 2 may be designated collecting phases and held at higher electrostatic potential relative to phase 3, which serves as a barrier phase to separate charge being collected in the high-field phases of the adjacent pixel. Following charge integration, transfer begins by holding only the phase-1 gates at high potential so that charge generated in that phase will collect there, and charge generated in the phase-2 and phase-3 phases, now both at zero potential, rapidly diffuses into the potential well under phase 1. Figure 3 illustrates the electrode structure defining each pixel of a three-phase CCD, and depicts electrons accumulating in the potential well underlying the phase-1 electrode, which is being held at a positive voltage (labeled +V). Charge transfer progresses with an appropriately timed sequence of voltages being applied to the gates in order to cause potential wells and barriers to migrate across each pixel.

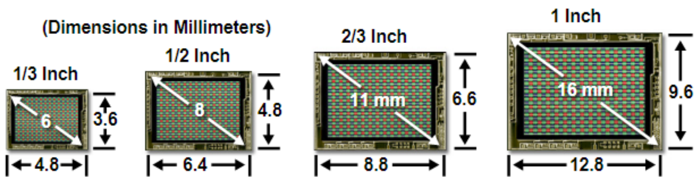


Figure 4. Common CCD image sensor formats.

At each transfer step, the voltage coupled to the well ahead of the charge packet is made positive while the electron-containing well is made negative or set to zero (ground), forcing the accumulated electrons to advance to the next phase. Rather than utilizing abrupt voltage transitions in the clocking sequence, the applied voltage changes on adjacent phases are gradual and overlap in order to ensure the most efficient charge transfer. The transition to phase 2 is carried out by applying positive potential to the phase-2 gates, spreading the collected charge between the phase-1 and phase-2 wells, and when the phase-1 potential is returned to ground, the entire charge packet is forced into phase 2. A similar sequence of timed voltage transitions, under control of the parallel shift register clock, is employed to shift the charge from phase 2 to phase 3, and the process continues until an entire single-pixel shift has been completed. One three-phase clock cycle applied to the entire parallel register results in a single-row shift of the entire array. An important factor in three-phase transfer is that a potential barrier is always maintained between adjacent pixel charge packets, which allows the one-to-one spatial correspondence between sensor and display pixels to be maintained throughout the image capture sequence.

Figure 6 illustrates the sequence of operations just described for charge transfer in a three-phase CCD, as well as the clocking sequence for drive pulses supplied by the parallel shift register clock to accomplish the transfer. In this schematic visualization of the pixel, charge is depicted being transferred from left to right by clocking signals that simultaneously decrease the voltage on the positively-biased electrode (defining a potential well) and increase it on the electrode to the right. In the last of the three steps, charge has been completely transferred from one gate electrode to the next. Note that the rising and falling phases of the clock drive pulses are timed to overlap slightly (not illustrated) in order to more efficiently transfer charge and to minimize the possibility of charge loss during the shift.

With each complete parallel transfer, charge packets from an entire pixel row are moved into the serial register where they can be sequentially shifted toward the output amplifier, as illustrated in the bucket brigade analogy. This horizontal (serial) transfer utilizes the same three-phase charge-coupling mechanism as the vertical row-shift, with timing control provided in this case by signals from the serial shift register clock. After all pixels are transferred from the serial register for readout, the parallel register clock provides the time signals for shifting the next row of trapped photoelectrons into the serial register. Each charge packet in the serial register is delivered to the CCD's output node where it is detected and read by an output amplifier (sometimes referred to as the on-chip preamplifier) that converts the charge into a proportional voltage. The voltage output of the amplifier represents the signal magnitude produced by successive photodiodes, as read out in sequence from left to right in each row and from the top row to the bottom over the entire two-dimensional array. The CCD output at this stage is, therefore, an analog voltage signal equivalent to a raster scan of accumulated charge over the imaging surface of the device.

After the output amplifier fulfills its function of magnifying a charge packet and converting it to a proportional voltage, the signal is transmitted to an analog-to-digital converter (ADC), which converts the voltage value into the 0 and 1 binary code necessary for interpretation by the computer. Each pixel is assigned a digital value corresponding to signal amplitude, in steps sized according to the resolution, or bit depth, of the ADC. For example, an ADC capable of 12-bit resolution assigns each pixel a value ranging from 0 to 4095, representing 4096 possible image gray levels (2 to the 12th power is equal to 4096 digitizer steps). Each gray-level step is termed an analog-to-digital unit (ADU).

The technological sophistication of current CCD imaging systems is remarkable considering the large number of operations required to capture a digital image, and the accuracy and speed with which the process is accomplished. The sequence of events required to capture a single image with a full-frame CCD camera system can be summarized as follows:

- Camera shutter is opened to begin accumulation of photoelectrons, with the gate electrodes biased appropriately for charge collection.
- At the end of the integration period, the shutter is closed and accumulated charge in pixels is shifted row by row across the parallel register under control of clock signals from the camera electronics. Rows of charge packets are transferred in sequence from one edge of the parallel register into the serial shift register.
- Charge contents of pixels in the serial register are transferred one pixel at a time into an output node to be read by an on-chip amplifier, which boosts the electron signal and converts it into an analog voltage output.
- An ADC assigns a digital value for each pixel according to its voltage amplitude.
- Each pixel value is stored in computer memory or in a camera frame buffer.

- The serial readout process is repeated until all pixel rows of the parallel register are emptied, which is commonly 1000 or more rows for high-resolution cameras.
- The complete image file in memory, which may be several megabytes in size, is displayed in a suitable format on the computer monitor for visual evaluation.
- The CCD is cleared of residual charge prior to the next exposure by executing the full readout cycle except for the digitization step.

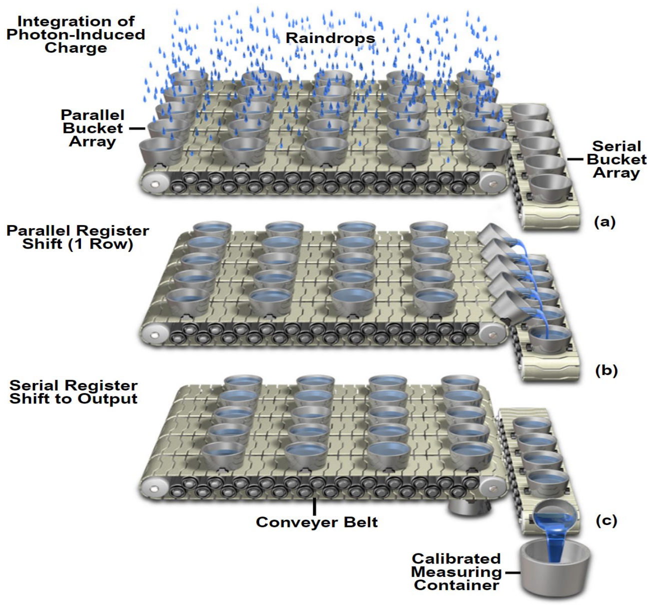


Figure 5. Bucket Brigade CCD Analogy.

In spite of the large number of operations performed, more than one million pixels can be transferred across the chip, assigned a gray-scale value with 12-bit resolution, stored in computer memory, and displayed in less than one second. A typical total time requirement for readout and image display is approximately 0.5 second for a 1-megapixel camera operating at a 5-MHz digitization rate. Charge transfer efficiency can also be extremely high for cooled-CCD cameras, with minimal loss of charge occurring, even with the thousands of transfers required for pixels in regions of the array that are farthest from the output amplifier.

8.1.3 CCD Image Sensor Architecture

Three basic variations of CCD architecture are in common use for imaging systems: full frame, frame transfer, and interline transfer. The full-frame CCD has the advantage of nearly 100-percent of its surface being photosensitive, with virtually no dead space between pixels. The imaging surface must be protected from incident light during readout of the CCD, and for this reason, an electromechanical shutter is usually employed for controlling exposures. Charge accumulated with the shutter open is subsequently transferred and read out after the shutter is closed, and because the two steps cannot occur simultaneously, image frame rates are limited by the mechanical shutter speed, the charge-transfer rate, and readout steps. Although full-frame devices have the largest photosensitive area of the CCD types, they are most useful with specimens having high intra-scene dynamic range, and in applications that do not require time resolution of less than approximately one second. When operated in a subarray mode (in which a reduced portion of the full pixel array is read out) in order to accelerate readout, the fastest frame rates possible are on the order of 10 frames per second, limited by the mechanical shutter.

Frame-transfer CCDs can operate at faster frame rates than full-frame devices because exposure and readout can occur simultaneously with various degrees of overlap in timing. They are similar to full-frame devices in structure of the parallel register, but one-half of the rectangular pixel array is covered by an opaque mask, and is used as a storage buffer for photoelectrons gathered by the unmasked light-sensitive portion. Following image exposure, charge accumulated in the photosensitive pixels is rapidly shifted to pixels on the storage side of the chip, typically within approximately 1 millisecond. Because the storage pixels are protected from light exposure by an aluminum or similar opaque coating, stored charge in that portion of the sensor can be systematically read out at a slower, more efficient rate while the next image is simultaneously being exposed on the photosensitive side of the chip. A camera shutter is not necessary because the time required for charge transfer from the image area to the storage area of the chip is only a fraction of the time needed for a typical exposure. Because cameras utilizing frame-transfer CCDs can be operated continuously at high frame rates without mechanical shuttering, they are suitable for investigating rapid kinetic processes by methods such as dye ratio imaging, in which high spatial resolution and dynamic range are important. A disadvantage of this sensor type is that only one-half of the surface area of the CCD is used for imaging, and consequently, a much larger chip is required than for a full-frame device with an equivalent-size imaging array, adding to the cost and imposing constraints on the physical camera design.

In the interline-transfer CCD design, columns of active imaging pixels and masked storage-transfer pixels alternate over the entire parallel register array. Because a charge-transfer channel is located immediately adjacent to each photosensitive pixel column, stored charge must only be shifted one column into a transfer channel. This single transfer step can be performed in less than 1 millisecond, after which the storage array is read out by a series of parallel shifts into the serial register while the image array is being exposed for the next image. The interline-transfer architecture allows very short integration periods through electronic control of exposure intervals, and in place of a mechanical shutter, the array can be rendered effectively light-insensitive by discarding accumulated charge rather than shifting it to the transfer channels. Although interline-transfer sensors allow video-rate readout and high-quality images of brightly illuminated subjects, basic forms of earlier devices suffered from reduced dynamic range, resolution, and sensitivity, due to the fact that approximately 75 percent of the CCD surface is occupied by the storage-transfer channels.

Although earlier interline-transfer CCDs, such as those used in video camcorders, offered high readout speed and rapid frame rates without the necessity of shutters, they did not provide adequate performance for low-light high-resolution applications in microscopy. In addition to the reduction in light-sensitivity attributable to the alternating columns of imaging and storage-transfer regions, rapid readout rates led to higher camera read noise and reduced dynamic range in earlier interline-transfer imagers. Improvements in sensor design and camera electronics have completely changed the situation to the extent that current interline devices provide superior performance for digital microscopy cameras, including those used for low-light applications such as recording small concentrations of fluorescent molecules. Adherent microlenses, aligned on the CCD surface to cover pairs of image and storage pixels, collect light that would normally be lost on the masked pixels and focus it on the light-sensitive pixels. By combining small pixel size with microlens technology, interline sensors are capable of delivering spatial resolution and light-collection efficiency comparable to full-frame and frame-transfer CCDs. The effective photosensitive area of interline sensors utilizing on-chip microlenses is increased to 75-90 percent of the surface area.

An additional benefit of incorporating microlenses in the CCD structure is that the spectral sensitivity of the sensor can be extended into the blue and ultraviolet wavelength regions, providing enhanced utility for shorter-wavelength applications, such as popular fluorescence techniques employing green fluorescent protein (GFP) and dyes excited by ultraviolet light. In order to increase quantum efficiency across the visible spectrum, recent high-performance chips incorporate gate structures composed of materials such as indium tin oxide, which have much higher transparency in the blue-green spectral region. Such nonabsorbing gate structures result in quantum efficiency values approaching 80 percent for green light.

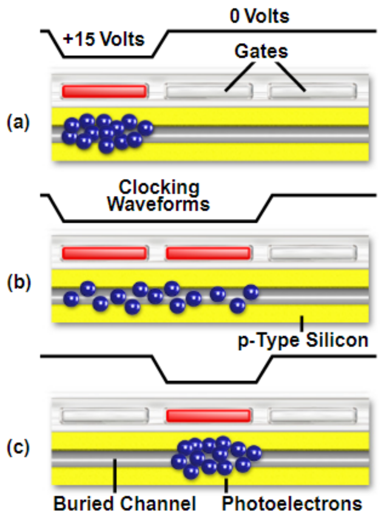


Figure 6. Three Phase CCD Clocking Systems.

The past limitation of reduced dynamic range for interline-transfer CCDs has largely been overcome by improved electronic technology that has lowered camera read noise by approximately one-half. Because the active pixel area of interline CCDs is approximately one-third that of comparable full-frame devices, the full well capacity (a function of pixel area) is similarly reduced. This factor, combined with relatively high camera read noise, resulted in insufficient signal dynamic range to support more than 8 or 10-bit digitization. High-performance interline cameras now operate with read noise values as low as 4 to 6 electrons, resulting in dynamic range performance equivalent to that of 12-bit cameras employing full-frame CCDs. Additional improvements in chip design factors such as clocking schemes, and in camera electronics, have enabled increased readout rates. Interline-transfer CCDs now enable 12-bit megapixel images to be acquired at 20-megahertz rates, approximately 4 times the rate of full-frame cameras with comparable array sizes. Other technological improvements, including modifications of the semiconductor composition, are incorporated in some interline-transfer CCDs to improve quantum efficiency in the near-infrared portion of the spectrum.

8.1.4 CCD Detector Imaging Performance

Several camera operation parameters that modify the readout stage of image acquisition have an impact on image quality. The readout rate of most scientific-grade CCD cameras is adjustable, and typically ranges from approximately 0.1 MHz to 10 or 20 MHz. The maximum achievable rate is a function of the processing speed of the ADC and other camera electronics, which reflect the time required to digitize a single pixel. Applications aimed at tracking rapid kinetic processes require fast readout and frame rates in order to achieve adequate temporal resolution, and in certain situations, a video rate of 30 frames per second or higher is necessary. Unfortunately, of the various noise components that are always present in an electronic image, read noise is a major source, and high readout rates increase the noise level. Whenever the highest temporal resolution is not required, better images of specimens that produce low pixel intensity values can be obtained at slower readout rates, which minimize noise and maintain adequate signal-to-noise ratio. When dynamic processes require rapid imaging frame rates, the normal CCD readout sequence can be modified to reduce the number of charge packets processed, enabling acquisition rates of hundreds of frames per second in some cases. This increased frame rate can be accomplished by combining pixels during CCD readout and/or by reading out only a portion of the detector array, as described below.

The image acquisition software of most CCD camera systems used in optical microscopy allows the user to define a smaller subset, or subarray, of the entire pixel array to be designated for image capture and display. By selecting a reduced portion of the image field for processing, unselected pixels are discarded without being digitized by the ADC, and readout speed is correspondingly increased. Depending upon the camera control software employed, a subarray may be chosen from pre-defined array sizes, or designated interactively as a region of interest using the computer mouse and the monitor display. The subarray readout technique is commonly utilized for acquiring sequences of time-lapse images, in order to produce smaller and more manageable image files.

Accumulated charge packets from adjacent pixels in the CCD array can be combined during readout to form a reduced number of superpixels. This process is referred to as pixel binning, and is performed in the parallel register by clocking two or more row shifts into the serial register prior to executing the serial shift and readout sequence. The binning process is usually repeated in the serial register by clocking multiple shifts into the readout node before the charge is read by the output amplifier. Any combination of parallel and serial shifts can be combined, but typically a symmetrical matrix of pixels are combined to form each single superpixel. As an example, 3 x 3 binning is accomplished by initially performing 3 parallel shifts of rows into the serial register (prior to serial transfer), at which point each pixel in the serial register contains the combined charge from 3 pixels, which were neighbors in adjacent parallel rows. Subsequently, 3 serial-shift steps are performed into the output node before the charge is measured. The resulting charge packet is processed as a single pixel, but contains the combined photoelectron content of 9 physical pixels (a 3 x 3 superpixel). Although binning reduces spatial resolution, the procedure often allows image acquisition under circumstances that make imaging impossible with normal CCD readout. It allows higher frame rates for image sequences if the acquisition rate is limited by the camera read cycle, as well as providing improved signal-to-noise ratio for equivalent exposure times. Additional advantages include shorter exposure times to produce the same image brightness (highly important for live cell imaging), and smaller image file sizes, which reduces computer storage demands and speeds image processing.

A third camera acquisition factor, which can affect image quality because it modifies the CCD readout process, is the electronic gain of the camera system. The gain adjustment of a digital CCD camera system defines the number of accumulated photoelectrons that determine each gray level step distinguished by the readout electronics, and is typically applied at the analog-to-digital conversion step. An increase in electronic gain corresponds to a decrease in the number of photoelectrons that are assigned per gray level (electrons/ADU), and allows a given signal level to be divided into a larger number of gray level steps. Note that this differs from gain adjustments applied to photomultiplier tubes or vidicon tubes, in which the varying signal is amplified by a fixed multiplication factor. Although electronic gain adjustment does provide a method to expand a limited signal amplitude to a desired large number of gray levels, if it is used excessively, the small number of electrons that distinguish adjacent gray levels can lead to digitization errors. High gain settings can result in noise due to the inaccurate digitization, which appears as graininess in the final image. If a reduction in exposure time is desired, an increase in electronic gain will allow maintaining a fixed large number of gray scale steps, in spite of the reduced signal level, providing that the applied gain does not produce excessive image deterioration. As an example of the effect of different gain factors applied to a constant signal level, an initial gain setting that assigns 8 electrons per ADU (gray level) dictates that a pixel signal consisting of 8000 electrons will be displayed at 1000 gray levels. By increasing the gain through application of a 4x gain factor to the base setting, the number of electrons per gray level is reduced to 2 (2 electrons/ADU), and 4000 gray levels are distinguished by the digitizing electronics.

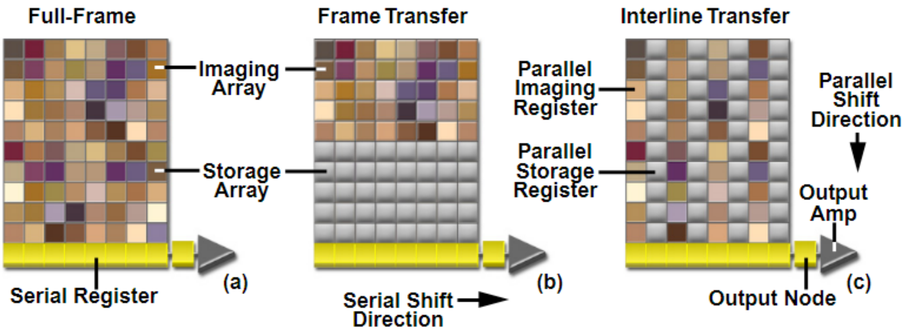


Figure 7. Common charge-coupled device (CCD) architectures.

Digital image quality can be assessed in terms of four quantifiable criteria that are determined in part by the CCD design, but which also reflect the implementation of the camera operation variables that directly affect the imaging performance of the CCD detector. The principal image quality criteria and their effects are summarized as follows:

- *Spatial Resolution*: Determines the ability to capture fine specimen details without pixels being visible in the image.
- *Light-Intensity Resolution*: Defines the dynamic range or number of gray levels that are distinguishable in the displayed image.
- *Time Resolution*: The sampling (frame) rate determines the ability to follow live specimen movement or rapid kinetic processes.
- *Signal-to-Noise Ratio*: Determines the visibility and clarity of specimen signals relative to the image background.

In microscope imaging, it is common that not all important image quality criteria can be simultaneously optimized in a single image, or image sequence. Obtaining the best images within the constraints imposed by a particular specimen or experiment typically requires a compromise among the criteria listed, which often exert contradictory demands. For example, capturing time-lapse sequences of live fluorescently-labeled specimens may require reducing the total exposure time to minimize photo bleaching and phototoxicity. Several methods can be utilized to accomplish this, although each involves a degradation of some aspect of imaging performance. If the specimen is exposed less frequently, temporal resolution is reduced; applying pixel binning to allow shorter exposures reduces spatial resolution; and increasing electronic gain compromises dynamic range and signal-to-noise ratio. Different situations often require completely different imaging rationales for optimum results. In order to maximize dynamic range in a single image of a specimen that requires a short exposure time, the application of binning or a gain increase may accomplish the goal without significant negative effects on the image. Performing efficient digital imaging requires the microscopist to be completely familiar with the crucial image quality criteria, and the practical aspects of balancing camera acquisition parameters to maximize the most significant factors in a particular situation.

8.1.5 CCD Camera Noise Sources

Camera sensitivity, in terms of the minimum detectable signal, is determined by both the photon statistical (shot) noise and electronic noise arising in the CCD. A conservative estimation is that a signal can only be discriminated from accompanying noise if it exceeds the noise by a factor of approximately 2.7 (SNR of 2.7). The minimum signal that can theoretically yield a given SNR value is determined by random variations of the photon flux, an inherent noise source associated with the signal, even with an ideal noiseless detector. This photon statistical noise is equal to the square root of the number of signal photons, and since it cannot be eliminated, it determines the maximum achievable SNR for a noise-free detector. The signal/noise ratio is therefore given by the signal level, S, divided by the square-root of the signal (S(1/2)), and is equal to the square-root of S. If a SNR value of 2.7 is required for discriminating signal from noise, a signal level of 8 photons is the minimum theoretically detectable light flux.

In practice, other noise components, which are not associated with the specimen photon signal, are contributed by the CCD and camera system electronics, and add to the inherent photon statistical noise. Once accumulated in collection wells, charge arising from noise sources cannot be distinguished from photon-derived signal. Most of the system noise results from readout amplifier noise and thermal electron generation in the silicon of the detector chip. The thermal noise is attributable to kinetic vibrations of silicon atoms in the CCD substrate that liberate electrons or holes even when the device is in total darkness, and which subsequently accumulate in the potential wells. For this reason, the noise is referred to as dark noise, and represents the uncertainty in the magnitude of dark charge accumulation during a specified time interval. The rate of generation of dark charge, termed dark current, is unrelated to photon-induced signal but is highly temperature dependent. In similarity to photon noise, dark noise follows a statistical (square-root) relationship to dark current, and therefore it cannot simply be subtracted from the signal. Cooling the CCD reduces dark charge accumulation by an order of magnitude for every 20-degree Celsius temperature decrease, and high-performance cameras are usually cooled during use. Cooling even to 0 degrees is highly advantageous, and at -30 degrees, dark noise is reduced to a negligible value for nearly any microscopy application.

Providing that the CCD is cooled, the remaining major electronic noise component is read noise, primarily originating with the on-chip preamplifier during the process of converting charge carriers into a voltage signal. Although the read noise is added uniformly to every pixel of the detector, its magnitude cannot be precisely determined, but only approximated by an average value, in units of electrons (root-mean-square or rms) per pixel. Some types of readout amplifier noise are frequency dependent, and in general, read noise increases with the speed of measurement of the charge in each pixel. The increase in noise at high readout and frame rates is partially a result of the greater amplifier bandwidth required at higher pixel clock rates. Cooling the CCD reduces the readout amplifier noise to some extent, although not to an insignificant level. A number of design enhancements are incorporated in current high-performance camera systems that greatly reduce the significance of read noise, however. One strategy for achieving high readout and frame rates without increasing noise is to electrically divide the CCD into two or more segments in order to shift charge in the parallel register toward multiple output amplifiers located at opposite edges or corners of the chip. This procedure allows charge to be read out from the array at a greater overall speed without excessively increasing the read rate (and noise) of the individual amplifiers.

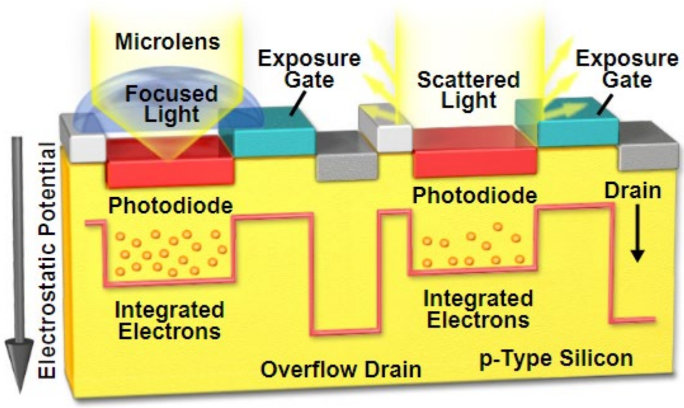


Figure 8. Microlens Array Interline CCD Technology.

Cooling the CCD in order to reduce dark noise provides the additional advantage of improving the charge transfer efficiency (CTE) of the device. This performance factor has become increasingly important due to the large pixel-array sizes employed in many current CCD imagers, as well as the faster readout rates required for investigations of rapid dynamic processes. With each shift of a charge packet along the transfer channels during the CCD readout process, a small portion may be left behind. While individual transfer losses at each pixel are miniscule in most cases, the large number of transfers required, especially in megapixel sensors, can result in significant losses for pixels at the greatest distance from the CCD readout amplifier(s) unless the charge transfer efficiency is extremely high. The occurrence of incomplete charge transfer can lead to image blurring due to the intermixing of charge from adjacent pixels. In addition, cumulative charge loss at each pixel transfer, particularly with large arrays, can result in the phenomenon of image shading, in which regions of images farthest away from the CCD output amplifier appear dimmer than those adjacent to the serial register. Charge transfer efficiency values in cooled CCDs can be 0.9999 or greater, and while CTEs this high are usually negligible in image effect, values lower than 0.999 are likely to produce shading.

Both hardware and software methods are available to compensate for image intensity shading. A software correction is implemented by capturing an image of a uniform-intensity field, which is then utilized by the imaging system to generate a pixel-by-pixel correction map that can be applied to subsequent specimen images to eliminate nonuniformity due to shading.

Software correction techniques are generally satisfactory in systems that do not require correction factors greater than approximately 10-20 percent of the local intensity. Larger corrections, up to approximately fivefold, can be handled by hardware methods through the adjustment of gain factors for individual pixel rows. The required gain adjustment is determined by sampling signal intensities in five or six masked reference pixels located outside the image area at the end of each pixel row. Voltage values obtained from the columns of reference pixels at the parallel register edge serve as controls for charge transfer loss, and produce correction factors for each pixel row that are applied to voltages obtained from that row during readout. Correction factors are large in regions of some sensors, such as areas distant from the output amplifier in video-rate cameras, and noise levels may be substantially increased for these image areas. Although the hardware correction process removes shading effects without apparent signal reduction, it should be realized that the resulting signal-to-noise ratio is not uniform over the entire image.

8.1.6 Spatial and Temporal Resolution in CCD Image Sensors

In many applications, an image capture system capable of providing high temporal resolution is a primary requirement. For example, if the kinetics of a process being studied necessitates video-rate imaging at moderate resolution, a camera capable of delivering superb resolution is, nevertheless, of no benefit if it only provides that performance at slow-scan rates, and performs marginally or not at all at high frame rates. Full-frame slow-scan cameras do not deliver high resolution at video rates, requiring approximately one second per frame for a large pixel array, depending upon the digitization rate of the electronics. If specimen signal brightness is sufficiently high to allow short exposure times (on the order of 10 milliseconds), the use of binning and subarray selection makes it possible to acquire about 10 frames per second at reduced resolution and frame size with cameras having electromechanical shutters. Faster frame rates generally necessitate the use of interline-transfer or frame-transfer cameras, which do not require shutters and typically can also operate at higher digitization rates. The latest generation of high-performance cameras of this design can capture full-frame 12-bit images at near video rates.

The now-excellent spatial resolution of CCD imaging systems is coupled directly to pixel size, and has improved consistently due to technological improvements that have allowed CCD pixels to be made increasingly smaller while maintaining other performance characteristics of the imagers. In comparison to typical film grain sizes (approximately 10 micrometers), the pixels of many CCD cameras employed in biological microscopy are smaller and provide more than adequate resolution when coupled with commonly used high-magnification objectives that project relatively large-radii diffraction (Airy) disks onto the CCD surface. Interline-transfer scientific-grade CCD cameras are now available having pixels smaller than 5 micrometers, making them suitable for high-resolution imaging even with low-magnification objectives. The relationship of detector element size to relevant optical resolution criteria is an important consideration in choosing a digital camera if the spatial resolution of the optical system is to be maintained.

The Nyquist sampling criterion is commonly utilized to determine the adequacy of detector pixel size with regard to the resolution capabilities of the microscope optics. Nyquist's theorem specifies that the smallest diffraction disk radius produced by the optical system must be sampled by at least two pixels in the imaging array in order to preserve the optical resolution and avoid aliasing. As an example, consider a CCD having pixel dimensions of 6.8 x 6.8 micrometers, coupled with a 100x, 1.3 numerical aperture objective, which produces a 26-micrometer (radius) diffraction spot at the plane of the detector. Excellent resolution is possible with this detector-objective combination, because the diffraction disk radius covers approximately a 4-pixel span (26 / 6.8 = 3.8 pixels) on the detector array, or nearly twice the Nyquist limiting criterion. At this sampling frequency, sufficient margin is available that the Nyquist criterion is nearly satisfied even with 2 x 2 pixel binning.

8.1.7 Image Sensor Quantum Efficiency

Detector quantum efficiency (QE) is a measure of the likelihood that a photon having a particular wavelength will be captured in the active region of the device to enable liberation of charge carriers. The parameter represents the effectiveness of a CCD imager in generating charge from incident photons, and is therefore a major determinant of the minimum detectable signal for a camera system, particularly when performing low-light-level imaging. No charge is generated if a photon never reaches the semiconductor depletion layer or if it passes completely through without transfer of significant energy. The nature of interaction between a photon and the detector depends upon the photon's energy and corresponding wavelength, and is directly related to the detector's spectral sensitivity range. Although conventional front-illuminated CCD detectors are highly sensitive and efficient, none have 100-percent quantum efficiencies at any wavelength.

Image sensors typically employed in fluorescence microscopy can detect photons within the spectral range of 400-1100 nanometers, with peak sensitivity normally in the range of 550-800 nanometers. Maximum QE values are only about 40-50 percent, except in the newest designs, which may reach 80 percent efficiency. Figure 10 illustrates the spectral sensitivity of a number of popular CCDs in a graph that plots quantum efficiency as a function of incident light wavelength. Most CCDs used in scientific imaging are of the interline-transfer type, and because the interline mask severely limits the photosensitive surface area, many older versions exhibit very low QE values. With the advent of the surface microlens technology to direct more incident light to the photosensitive regions between transfer channels, newer interline sensors are much more efficient and many have quantum efficiency values of 60-70 percent.

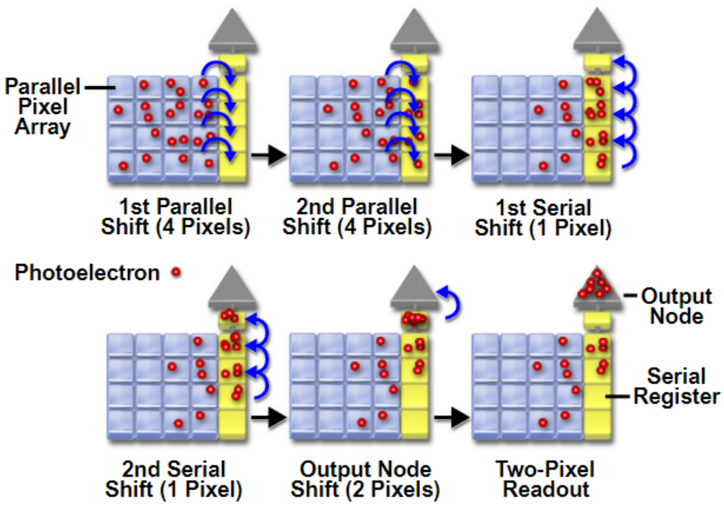


Figure 9. 2 x 2 Pixel Binning Electron Transfer Sequence.

Sensor spectral range and quantum efficiency are further enhanced in the ultraviolet, visible, and near-infrared wavelength regions by various additional design strategies in several high-performance CCDs. Because aluminum surface transfer gates absorb or reflect much of the blue and ultraviolet wavelengths, many newer designs employ other materials, such as indium-tin oxide, to improve transmission and quantum efficiency over a broader spectral range. Even higher QE values can be obtained with specialized back-thinned CCDs, which are constructed to allow illumination from the rear side, avoiding the surface electrode structure entirely. To make this possible, most of the silicon substrate is removed by etching, and although the resulting device is delicate and relatively expensive, quantum efficiencies of approximately 90 percent can routinely be achieved.

Other surface treatments and construction materials may be utilized to gain additional spectral-range benefits. Performance of back-thinned CCDs in the ultraviolet wavelength region is enhanced by the application of specialized antireflection coatings. Modified semiconductor materials are used in some detectors to improve quantum efficiency in the near-infrared. Sensitivity to wavelengths outside the normal spectral range of conventional front-illuminated CCDs can be achieved by the application of wavelength-conversion phosphors to the detector face. Phosphors for this purpose are chosen to absorb photon energy in the spectral region of interest and emit light within the spectral sensitivity region of the CCD. As an example of this strategy, if a specimen or fluorophore of interest emits light at 300 nanometers (where sensitivity of any CCD is minimal), a conversion phosphor can be employed on the detector surface that absorbs efficiently at 300 nanometers and emits at 560 nanometers, within the peak sensitivity range of the CCD.

8.1.8 Dynamic Range

A term referred to as the dynamic range of a CCD detector expresses the maximum signal intensity variation that can be quantified by the sensor. The quantity is specified numerically by most CCD camera manufacturers as the ratio of pixel full well capacity (FWC) to the read noise, with the rationale that this value represents the limiting condition in which intrascene brightness ranges from regions that are just at pixel saturation level to regions that are barely lost in noise. The sensor dynamic range determines the maximum number of resolvable gray-level steps into which the detected signal can be divided. To take full advantage of a CCD's dynamic range, it is appropriate to match the analog-to-digital converter's bit depth to the dynamic range in order to allow discrimination of as many gray scale steps as possible. For example, a camera with a 16,000-electron FWC and readout noise of 10 electrons, has a dynamic range of 1600, which supports between 10 and 11-bit A/D conversion. Analog-to-digital converters with bit depths of 10 and 11 are capable of discriminating

1024 and 2048 gray levels, respectively. Because a computer bit can only assume one of two possible states, the number of intensity steps that can be encoded by a digital processor (ADC) reflects its resolution (bit depth), and is equal to 2 raised to the value of the bit depth specification. Therefore, 8, 10, 12, and 14-bit processors can encode a maximum of 256, 1024, 4096, or 16384 gray levels.

Specifying dynamic range as the ratio of full well capacity to read noise is not necessarily a realistic measure of useful dynamic range, but is valuable for comparing sensors. In practice, useful dynamic range is smaller both because CCD response becomes nonlinear before full well capacity is reached and because a signal level equal to read noise is unacceptable visually and virtually useless for quantitative purposes. Note that the maximum dynamic range is not equivalent to the maximum possible signal-to-noise ratio, although the SNR is also a function of full well capacity. The photon statistical noise associated with the maximum possible signal, or FWC, is the square root of the FWC value, or 126 electrons. The maximum signal-to-noise ratio is therefore equal to the maximum signal divided by noise (16,000/126), or 126, the square root of the signal itself. The photon noise represents the minimum intrinsic noise level, and both detected stray light and electronic (system) noise diminish the maximum SNR that can be realized in practice to values below 126, since these sources reduce the effective FWC by adding charge that is not signal to the wells.

Although a manufacturer might typically equip a camera having a dynamic range of approximately 4000, for example, with a 12-bit ADC (4096 digitization steps), several factors are relevant in considering the match between sensor dynamic range and the digitizing capacity of the processor. For some of the latest interline-transfer CCD cameras that provide 12-bit digitization, the dynamic range determined from the FWC and read noise is approximately 2000, which would not normally require 12-bit processing. However, a number of the current designs include an option for setting gain at 0.5x, allowing full utilization of 12-bit resolution. This strategy takes advantage of the fact that pixels of the serial register are designed to have twice the electron capacity of parallel register pixels, and when the camera is operated in 2 x 2 binning mode (common in fluorescence microscopy), 12-bit high-quality images can be obtained.

It is important to be aware of the various mechanisms in which electronic gain can be manipulated to utilize the available bit depth of the processor, and when dynamic range of different cameras is being compared, the best approach is to calculate the value from the pixel full well capacity and camera read noise. It is common to see camera systems equipped with processing electronics having much higher digitizing resolution than required by the inherent dynamic range of the camera. In such a system, operation at the conventional 1x electronic gain setting results in a potentially large number of unused processor gray-scale levels. It is possible for the camera manufacturer to apply an unspecified gain factor of 2-4x, which might not be obvious to the user, and although this practice does boost the signal to utilize the full bit depth of the ADC, it produces increased digitization noise as the number of electrons constituting each gray level step is reduced.

The need for high bit depth in CCD cameras might be questioned in view of the fact that display devices such as computer monitors and many printers utilize only 8-bit processing, providing 256 gray levels, and other printed media as well as the human eye may only provide 5-7 bit discrimination. In spite of such low visual requirements, high bit-depth, high dynamic range camera systems are always advantageous, and are necessary for certain applications, particularly in fluorescence microscopy. When processing ratiometric or kinetic imaging data in quantitative investigations, larger numbers of gray levels allow light intensities to be more accurately determined. Additionally, when multiple image-processing operations are being performed, image data that are more precisely resolved into many gray level steps can withstand a greater degree of mathematical manipulation without exhibiting degradation as a result of arithmetic rounding-off errors.

A third advantage of high-bit imaging systems is realized when a portion of a captured image is selected for display, and the region of interest spans only a portion of the full image dynamic range. To optimize representation of the limited dynamic range, the original number of gray levels is typically expanded to occupy all 256 levels of an 8-bit monitor or print. Higher camera bit depth results in a less extreme expansion, and correspondingly less image degradation. As an example, if a selected image region spans only 5 percent of the full intrascene dynamic range, this represents over 200 gray levels of the 4096 discriminated by a 12-bit processor, but only 12 steps with an 8-bit (256 levels) system. When displayed at 256 levels on a monitor, or printed, the 12-level image expanded to this extent would appear pixelated, and exhibit blocky or contoured brightness steps rather than smooth tonal gradations.

8.1.9 Color CCD Image Sensors

Although CCDs are not inherently color sensitive, three different strategies are commonly employed to produce color images with CCD camera systems in order to capture the visual appearance of specimens in the microscope. Earlier technical difficulties in displaying and printing color images are no longer an issue, and the increase in information provided by color can be substantial. Many applications, such as fluorescence microscopy, the study of stained histology and pathology tissue sections, and other labeled specimen observations using brightfield or differential interference contrast techniques rely on color as an essential image component. The acquisition of color images with a CCD camera requires that red, green, and blue wavelengths be isolated by color filters, acquired separately, and subsequently combined into a composite color image.

Each approach utilized to achieve color discrimination has strengths and weaknesses, and all impose constraints that limit speed, lower temporal and spatial resolution, reduce dynamic range, and increase noise in color cameras compared to gray-scale cameras. The most common method is to blanket the CCD pixel array with an alternating mask of red, green, and blue (RGB) microlens filters arranged in a specific pattern, usually the Bayer mosaic pattern. Alternatively, with a three-chip design, the image is divided with a beam-splitting prism and color filters into three (RGB) components, which are captured with separate CCDs, and their outputs recombined into a color image. The third approach is a frame-sequential method that employs a single CCD to sequentially capture a separate image for each color by switching color filters placed in the illumination path or in front of the imager.

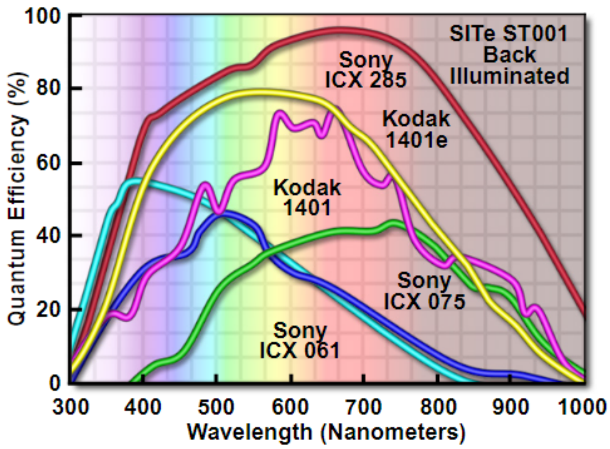


Figure 10. Scientific CCD Spectral Sensitivities.

The single-chip CCD with an adherent color filter array is used in most color microscopy cameras. The filter array consists of red, green, and blue microlenses applied over individual pixels in a regular pattern. The Bayer mosaic filter distributes color information over four-pixel sensor units that include one red, one blue, and two green filters. Green is emphasized in the distribution pattern to better conform to human visual sensitivity, and dividing color information among groups of four pixels only modestly degrades resolution. The human visual system acquires spatial detail primarily from the luminance component of color signals, and this information is retained in each pixel regardless of color. Visually satisfying images are achieved by combining color information of lower spatial resolution with the high-resolution monochrome structural details.

A unique design of single-CCD color cameras improves spatial resolution by slightly shifting the CCD between images taken in sequence, and then interpolating among them (a technique known as pixel-shifting), although image acquisition is slowed considerably by this process. Another approach to individual pixel masking is to rapidly move an array of color microlenses in a square pattern immediately over the CCD surface during photon collection. Finally, a recently introduced technology combines three photoelectron wells into each pixel at different depths for discrimination of photon wavelength. Maximum spatial resolution is retained in these strategies because each pixel provides red, green, and blue color information.

The three-chip color camera combines high spatial resolution with fast image acquisition, allowing high frame rates suitable for rapid image sequences and video output. By employing a beam splitter to direct signal to three filtered CCDs, which separately record red, green, and blue image components simultaneously, very high capture speeds are possible. However, because the light intensity delivered to each CCD is substantially reduced, the combined color image is much dimmer than a monochrome single-chip image given a comparable exposure. Gain may be applied to the color image to increase its brightness, but signal-to-noise ratio suffers, and images exhibit greater apparent noise. Spatial resolution attained by three-chip cameras can be higher than that of the individual CCD sensors if each CCD is offset by a subpixel amount relative to the others. Since the red, green, and blue images represent slightly different samples, they can be combined by the camera software to produce higher-resolution composite images. Many microscopy and other scientific applications that require high spatial and temporal resolution benefit from the use of triple-CCD camera systems.

Color cameras referred to as frame-sequential are equipped with a motorized filter wheel or liquid-crystal tunable filter (LCTF) to sequentially expose red, green, and blue component images onto a single CCD. Because the same sensor is used for separate red, green, and blue images, the full spatial resolution of the chip is maintained, and image registration is automatically obtained. The acquisition of three frames in succession slows the process of image acquisition and display, and proper color balance often requires different integration times for the three colors. Although this camera type is not generally appropriate for high-frame-rate acquisition, the use of rapidly-responding liquid-crystal tunable filters for the R-G-B sequencing can increase the operation speed substantially. The polarization sensitivity of LCTFs must be considered in some applications, as they only transmit one polarization vector, and may alter the colors of birefringent specimens viewed in polarized light.

8.2 HISTORY

The basis for the CCD is the metal–oxide–semiconductor (MOS) structure, with MOS capacitors being the basic building blocks of a CCD, and a depleted MOS structure used as the photodetector in early CCD devices.

In the late 1960s, Willard Boyle and George E. Smith at Bell Labs were researching MOS technology while working on semiconductor bubble memory. They realized that an electric charge was the analogy of the magnetic bubble and that it could be stored on a tiny MOS capacitor. As it was fairly straightforward to fabricate a series of MOS capacitors in a row, they connected a suitable voltage to them so that the charge could be stepped along from one to the next. This led to the invention of the charge-coupled device by Boyle and Smith in 1969. They conceived of the design of what they termed, in their notebook, "Charge 'Bubble' Devices".

The initial paper describing the concept in April 1970 listed possible uses as memory, a delay line, and an imaging device. The device could also be used as a shift register. The essence of the design was the ability to transfer charge along the surface of a semiconductor from one storage capacitor to the next. The concept was similar in principle to the bucket-brigade device (BBD), which was developed at Philips Research Labs during the late 1960s.

The first experimental device demonstrating the principle was a row of closely spaced metal squares on an oxidized silicon surface electrically accessed by wire bonds. It was demonstrated by Gil Amelio, Michael Francis Tompsett and George Smith in April 1970. This was the first experimental application of the CCD in image sensor technology, and used a depleted MOS structure as the photodetector. The first patent on the application of CCDs to imaging was assigned to Tompsett, who filed the application in 1971.

The first working CCD made with integrated circuit technology was a simple 8-bit shift register, reported by Tompsett, Amelio and Smith in August 1970. This device had input and output circuits and was used to demonstrate its use as a shift register and as a crude eight pixel linear imaging device. Development of the device progressed at a rapid rate. By 1971, Bell researchers led by Michael Tompsett were able to capture images with simple linear devices. Several companies, including Fairchild Semiconductor, RCA and Texas Instruments, picked up on the invention and began development programs. Fairchild's effort, led by ex-Bell researcher Gil Amelio, was the first with commercial devices, and by 1974 had a linear 500-element device and a 2-D 100 × 100 pixel device. Steven Sasson, an electrical engineer working for Kodak, invented the first digital still camera using a Fairchild 100 × 100 CCD in 1975.

The interline transfer (ILT) CCD device was proposed by L. Walsh and R. Dyck at Fairchild in 1973 to reduce smear and eliminate a mechanical shutter. To further reduce smear from bright light sources, the frame-interline-transfer (FIT) CCD architecture was developed by K. Horii, T. Kuroda and T. Kunii at Matsushita (now Panasonic) in 1981.

The first KH-11 KENNEN reconnaissance satellite equipped with charge-coupled device array (800 × 800 pixels) technology for imaging was launched in December 1976. Under the leadership of Kazuo Iwama, Sony started a large development effort on CCDs involving a significant investment. Eventually, Sony managed to mass-produce CCDs for their camcorders. Before this happened, Iwama died in August 1982; subsequently, a CCD chip was placed on his tombstone to acknowledge his contribution. The first mass-produced consumer CCD video camera, the CCD-G5, was released by Sony in 1983, based on a prototype developed by Yoshiaki Hagiwara in 1981.

Early CCD sensors suffered from shutter lag. This was largely resolved with the invention of the pinned photodiode (PPD). It was invented by Nobukazu Teranishi, Hiromitsu Shiraki and Yasuo Ishihara at NEC in 1980. They recognized that lag can be eliminated if the signal carriers could be transferred from the photodiode to the CCD. This led to their invention of the pinned photodiode, a photodetector structure with low lag, low noise, high quantum efficiency and low dark current. It was first publicly reported by Teranishi and Ishihara with A. Kohono, E. Oda and

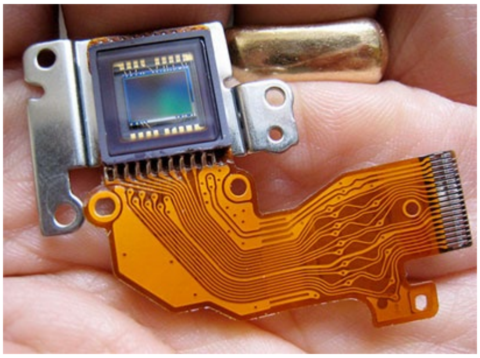
K. Arai in 1982, with the addition of an anti-blooming structure. The new photodetector structure invented at NEC was given the name "pinned photodiode" (PPD) by B.C. Burkey at Kodak in 1984. In 1987, the PPD began to be incorporated into most CCD devices, becoming a fixture in consumer electronic video cameras and then digital still cameras. Since then, the PPD has been used in nearly all CCD sensors and then CMOS sensors.

In January 2006, Boyle and Smith were awarded the National Academy of Engineering Charles Stark Draper Prize, and in 2009 they were awarded the Nobel Prize for Physics, for their invention of the CCD concept. Michael Tompsett was awarded the 2010 National Medal of Technology and Innovation, for pioneering work and electronic technologies including the design and development of the first CCD imagers. He was also awarded the 2012 IEEE Edison Medal for "pioneering contributions to imaging devices including CCD Imagers, cameras and thermal imagers".

8.3 BASICS OF OPERATION

In a CCD for capturing images, there is a photoactive region (an epitaxial layer of silicon), and a transmission region made out of a shift register.

An image is projected through a lens onto the capacitor array (the photoactive region), causing each capacitor to accumulate an electric charge proportional to the light intensity at that location. A one-dimensional array, used in line-scan cameras, captures a single slice of the image, whereas a two-dimensional array, used in video and still cameras, captures a two-dimensional picture corresponding to the scene projected onto the focal plane of the sensor. Once the array has been exposed to the image, a control circuit causes each capacitor to transfer its contents to its neighbor (operating as a shift register). The last capacitor in the array dumps its charge into a charge amplifier, which converts the charge into a voltage. By repeating this process, the controlling circuit converts the entire contents of the array in the semiconductor to a sequence of voltages. In a digital device, these voltages are then sampled, digitized, and usually stored in memory; in an analog device (such as an analog video camera), they are processed into a continuous analog signal (e.g. by feeding the output of the charge amplifier into a low-pass filter), which is then processed and fed out to other circuits for transmission, recording, or other processing.



Detailed Physics of Operation

Charge generation

Before the MOS capacitors are exposed to light, they are biased into the depletion region; in n-channel CCDs, the silicon under the bias gate is slightly p-doped or intrinsic. The gate is then biased at a positive potential, above the threshold for strong inversion, which will eventually result in the creation of an n channel below the gate as in a MOSFET. However, it takes time to reach this thermal equilibrium: up to hours in high-end scientific cameras cooled at low temperature. Initially after biasing, the holes are pushed far into the substrate, and no mobile electrons are at or near the surface; the CCD thus operates in a non-equilibrium state called deep depletion. Then, when electron–hole pairs are generated in the depletion region, they are separated by the electric field, the electrons move toward the surface, and the holes move toward the substrate.

Four pair-generation processes can be identified:

- photo-generation (up to 95% of quantum efficiency),
- generation in the depletion region,
- generation at the surface, and
- generation in the neutral bulk.

The last three processes are known as dark-current generation, and add noise to the image; they can limit the total usable integration time. The accumulation of electrons at or near the surface can proceed either until image integration is over and charge begins to be transferred, or thermal equilibrium is reached. In this case, the well is said to be full. The maximum capacity of each well is known as the well depth, typically about 105 electrons per pixel.

Design and Manufacturing

The photoactive region of a CCD is, generally, an epitaxial layer of silicon. It is lightly p doped (usually with boron) and is grown upon a substrate material, often p++. In buried-channel devices, the type of design utilized in most modern CCDs, certain areas of the surface of the silicon are ion implanted with phosphorus, giving them an n-doped designation. This region defines the channel in which the photogenerated charge packets will travel.

Later in the process, polysilicon gates are deposited by chemical vapor deposition, patterned with photolithography, and etched in such a way that the separately phased gates lie perpendicular to the channels. The channels are further defined by utilization of the LOCOS process to produce the channel stop region.

Channel stops are thermally grown oxides that serve to isolate the charge packets in one column from those in another. These channel stops are produced before the polysilicon gates are, as the LOCOS process utilizes a high-temperature step that would destroy the gate material. The channel stops are parallel to, and exclusive of, the channel, or “charge carrying”, regions.

Channel stops often have a p+ doped region underlying them, providing a further barrier to the electrons in the charge packets (this discussion of the physics of CCD devices assumes an electron transfer device, though hole transfer is possible).

The clocking of the gates, alternately high and low, will forward and reverse bias the diode that is provided by the buried channel (n-doped) and the epitaxial layer (p-doped). This will cause the CCD to deplete, near the p–n junction and will collect and move the charge packets beneath the gates—and within the channels—of the device.

CCD manufacturing and operation can be optimized for different uses. The above process describes a frame transfer CCD. While CCDs may be manufactured on a heavily doped p++ wafer it is also possible to manufacture a device inside p-wells that have been placed on an n-wafer. This second method, reportedly, reduces smear, dark current, and infrared and red response. This method of manufacture is used in the construction of interline-transfer devices.

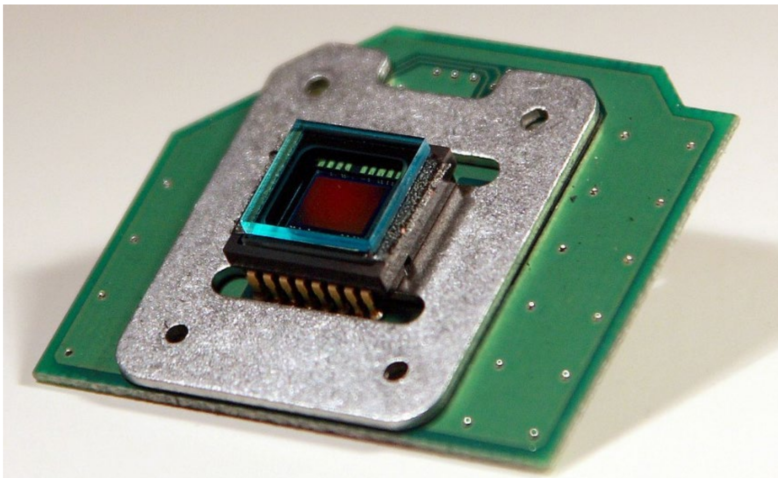
Another version of CCD is called a peristaltic CCD. In a peristaltic charge-coupled device, the charge-packet transfer operation is analogous to the peristaltic contraction and dilation of the digestive system. The peristaltic CCD has an additional implant that keeps the charge away from the silicon/silicon dioxide interface and generates a large lateral electric field from one gate to the next. This provides an additional driving force to aid in transfer of the charge packets.

8.4 ARCHITECTURE OF CHARGE-COUPLED DEVICE

The CCD image sensors can be implemented in several different architectures. The most common are full-frame, frame-transfer, and interline. The distinguishing characteristic of each of these architectures is their approach to the problem of shuttering.

In a full-frame device, all of the image area is active, and there is no electronic shutter. A mechanical shutter must be added to this type of sensor or the image smears as the device is clocked or read out.

With a frame-transfer CCD, half of the silicon area is covered by an opaque mask (typically aluminum). The image can be quickly transferred from the image area to the opaque area or storage region with acceptable smear of a few percent. That image can then be read out slowly from the storage region while a new image is integrating or exposing in the active area. Frame-transfer devices typically do not require a mechanical shutter and were a common architecture for early solid-state broadcast cameras. The downside to the frame-transfer architecture is that it requires twice the silicon real estate of an equivalent full-frame device; hence, it costs roughly twice as much.



The interline architecture extends this concept one step further and masks every other column of the image sensor for storage. In this device, only one pixel shift has to occur to transfer from image area to storage area; thus, shutter times can be less than a microsecond and smear is essentially eliminated. The advantage is not free, however, as the imaging area is now covered by opaque strips dropping the fill factor to approximately 50 percent and the effective quantum efficiency by an equivalent amount. Modern designs have addressed this deleterious characteristic by adding microlenses on the surface of the device to direct light away from the opaque regions and on the active area. Microlenses can bring the fill factor back up to 90 percent or more depending on pixel size and the overall system's optical design.

The choice of architecture comes down to one of utility. If the application cannot tolerate an expensive, failure-prone, power-intensive mechanical shutter, an interline device is the right choice. Consumer snap-shot cameras have used interline devices. On the other hand, for those applications that require the best possible light collection and issues of money, power and time are less important, the full-frame device is the right choice. Astronomers tend to prefer full-frame devices. The frame-transfer falls in between and was a common choice before the fill-factor issue of interline devices was addressed. Today, frame-transfer is usually chosen when an interline architecture is not available, such as in a back-illuminated device.

CCDs containing grids of pixels are used in digital cameras, optical scanners, and video cameras as light-sensing devices. They commonly respond to 70 percent of the incident light (meaning a quantum efficiency of about 70 percent) making them far more efficient than photographic film, which captures only about 2 percent of the incident light.

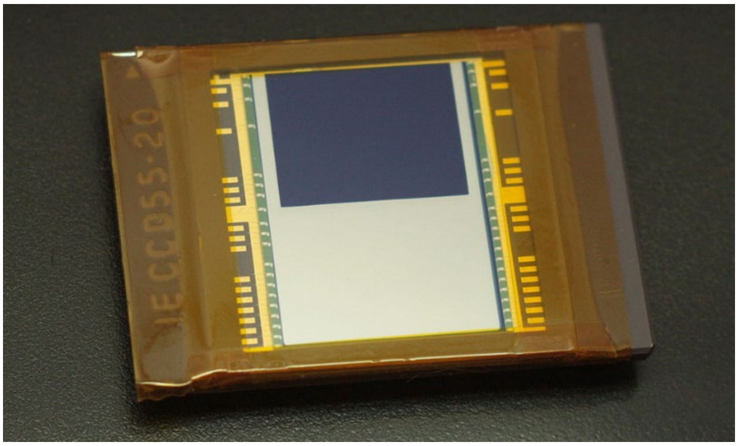
Most common types of CCDs are sensitive to near-infrared light, which allows infrared photography, night-vision devices, and zero lux (or near zero lux) video-recording/photography. For normal silicon-based detectors, the sensitivity is limited to 1.1 μm. One other consequence of their sensitivity to infrared is that infrared from remote controls often appears on CCD-based digital cameras or camcorders if they do not have infrared blockers.

Cooling reduces the array's dark current, improving the sensitivity of the CCD to low light intensities, even for ultraviolet and visible wavelengths. Professional observatories often cool their detectors with liquid nitrogen to reduce the dark current, and therefore the thermal noise, to negligible levels.

8.4.1 Frame transfer CCD

The frame transfer CCD imager was the first imaging structure proposed for CCD Imaging by Michael Tompsett at Bell Laboratories. A frame transfer CCD is a specialized CCD, often used in astronomy and some professional video cameras, designed for high exposure efficiency and correctness.

The normal functioning of a CCD, astronomical or otherwise, can be divided into two phases: exposure and readout. During the first phase, the CCD passively collects incoming photons, storing electrons in its cells. After the exposure time is passed, the cells are read out one line at a time. During the readout phase, cells are shifted down the entire area of the CCD. While they are shifted, they continue to collect light. Thus, if the shifting is not fast enough, errors can result from light that falls on a cell holding charge during the transfer. These errors are referred to as "vertical smear" and cause a strong light source to create a vertical line above and below its exact location. In addition, the CCD cannot be used to collect light while it is being read out. Unfortunately, a faster shifting requires a faster readout, and a faster readout can introduce errors in the cell charge measurement, leading to a higher noise level.



A frame transfer CCD solves both problems: it has a shielded, not light sensitive, area containing as many cells as the area exposed to light. Typically, this area is covered by a reflective material such as aluminum. When the exposure time is up, the cells are transferred very rapidly to the hidden area. Here, safe from any incoming light, cells can be read out at any speed one deems necessary to correctly measure the cells' charge. At the same time, the exposed part of the CCD is collecting light again, so no delay occurs between successive exposures.

The disadvantage of such a CCD is the higher cost: the cell area is basically doubled, and more complex control electronics are needed.

8.4.2 Intensified Charge-coupled Device

An intensified charge-coupled device (ICCD) is a CCD that is optically connected to an image intensifier that is mounted in front of the CCD.

An image intensifier includes three functional elements: a photocathode, a micro-channel plate (MCP) and a phosphor screen. These three elements are mounted one close behind the other in the mentioned sequence. The photons which are coming from the light source fall onto the photocathode, thereby generating photoelectrons. The photoelectrons are accelerated towards the MCP by an electrical control voltage, applied between photocathode and MCP. The electrons are multiplied inside of the MCP and thereafter accelerated towards the phosphor screen. The phosphor screen finally converts the multiplied electrons back to photons which are guided to the CCD by a fiber optic or a lens.

An image intensifier inherently includes a shutter functionality: If the control voltage between the photocathode and the MCP is reversed, the emitted photoelectrons are not accelerated towards the MCP but return to the photocathode. Thus, no electrons are multiplied and emitted by the MCP, no electrons are going to the phosphor screen and no light is emitted from the image intensifier. In this case no light falls onto the CCD, which means that the shutter is closed. The process of reversing the control voltage at the photocathode is called gating and therefore ICCDs are also called gateable CCD cameras.

Besides the extremely high sensitivity of ICCD cameras, which enable single photon detection, the gateability is one of the major advantages of the ICCD over the EMCCD cameras. The highest performing ICCD cameras enable shutter times as short as 200 picoseconds.

ICCD cameras are in general somewhat higher in price than EMCCD cameras because they need the expensive image intensifier. On the other hand, EMCCD cameras need a cooling system to cool the EMCCD chip down to temperatures around 170 K (−103 °C). This cooling system adds additional costs to the EMCCD camera and often yields heavy condensation problems in the application.

ICCDs are used in night vision devices and in various scientific applications.

8.4.3 Electron-multiplying CCD

An electron-multiplying CCD (EMCCD, also known as an L3Vision CCD, a product commercialized by e2v Ltd., GB, L3CCD or Impactron CCD, a now-discontinued product offered in the past by Texas Instruments) is a charge-coupled device in which a gain register is placed between the shift register and the output amplifier. The gain register is split up into a large number of stages. In each stage, the electrons are multiplied by impact ionization in a similar way to an avalanche diode. The gain probability at every stage of the register is small ($P < 2\%$), but as the number of elements is large ($N > 500$), the overall gain can be very high ($g = (1 + P)^N$), with single input electrons giving many thousands of output electrons. Reading a signal from a CCD gives a noise background, typically a few electrons. In an EMCCD, this noise is superimposed on many thousands of electrons rather than a single electron; the devices' primary advantage is thus their negligible readout noise. The use of avalanche breakdown for amplification of photo charges had already been described in the U.S. Patent 3,761,744 in 1973 by George E. Smith/Bell Telephone Laboratories.

EMCCDs show a similar sensitivity to intensified CCDs (ICCDs). However, as with ICCDs, the gain that is applied in the gain register is stochastic and the exact gain that has been applied to a pixel's charge is impossible to know. At high gains (> 30), this uncertainty has the same effect on the signal-to-noise ratio (SNR) as halving the quantum efficiency (QE) with respect to operation with a gain of unity. However, at very low light levels (where the quantum efficiency is most important), it can be assumed that a pixel either contains an electron—or not. This removes the noise associated with the stochastic multiplication at the risk of counting multiple electrons in the same pixel as a single electron. To avoid multiple counts in one pixel due to coincident photons in this mode of operation, high frame rates are essential. The dispersion in the gain is shown in the graph on the right.

For multiplication registers with many elements and large gains it is well modelled by the equation:

$$P(n) = \frac{(n-m+1)^{m-1}}{(m-1)!\left(g-1+\frac{1}{m}\right)^m} \exp\left(-\frac{n-m+1}{g-1+\frac{1}{m}}\right) \quad \text{if } n \geq m$$

where P is the probability of getting n output electrons given m input electrons and a total mean multiplication register gain of g.

Because of the lower costs and better resolution, EMCCDs are capable of replacing ICCDs in many applications. ICCDs still have the advantage that they can be gated very fast and thus are useful in applications like range-gated imaging. EMCCD cameras indispensably need a cooling system—using either thermoelectric cooling or liquid nitrogen—to cool the chip down to temperatures in the range of −65 to −95 °C (−85 to −139 °F). This cooling system unfortunately adds additional costs to the EMCCD imaging system and may yield condensation problems in the application. However, high-end EMCCD cameras are equipped with a permanent hermetic vacuum system confining the chip to avoid condensation issues.

The low-light capabilities of EMCCDs find use in astronomy and biomedical research, among other fields. In particular, their low noise at high readout speeds makes them very useful for a variety of astronomical applications involving low light sources and transient events such as lucky imaging of faint stars, high speed photon counting photometry, Fabry-Pérot spectroscopy and high-resolution spectroscopy. More recently, these types of CCDs have broken into the field of biomedical research in low-light applications including small animal imaging, single-molecule imaging, Raman spectroscopy, super resolution microscopy as well as a wide variety of modern fluorescence microscopy techniques thanks to greater SNR in low-light conditions in comparison with traditional CCDs and ICCDs.

In terms of noise, commercial EMCCD cameras typically have clock-induced charge (CIC) and dark current (dependent on the extent of cooling) that together lead to an effective readout noise ranging from 0.01 to 1 electrons per pixel read. However, recent improvements in EMCCD technology have led to a new generation of cameras capable of producing significantly less CIC, higher charge transfer efficiency and an EM gain 5 times higher than what was previously available. These advances in low-light detection lead to an effective total background noise of 0.001 electrons per pixel read, a noise floor unmatched by any other low-light imaging device.

SUMMARY

- A charge-coupled device (CCD) is an integrated circuit containing an array of linked, or coupled, capacitors.
- A charge coupled device (CCD) is a highly sensitive photon detector. The CCD is divided up into a large number of light-sensitive small areas (known as pixels) which can be used to build up an image of the scene of interest.
- Digital camera systems, incorporating a variety of charge-coupled device (CCD) detector configurations, are by far the most common image capture technology employed in modern optical microscopy.
- Digital cameras replace the sensitized film with a CCD photon detector, a thin silicon wafer divided into a geometrically regular array of thousands or millions of light-sensitive regions that capture and store image information in the form of localized electrical charge that varies with incident light intensity.
- In general, the stored charge is linearly proportional to the light flux incident on a sensor pixel up to the capacity of the well; consequently this full-well capacity (FWC) determines the maximum signal that can be sensed in the pixel, and is a primary factor affecting the CCD's dynamic range.
- The general clocking scheme employed in three-phase transfer begins with a charge integration step, in which two of the three parallel phases per pixel are set to a high bias value, producing a high-field region relative to the third gate, which is held at low or zero potential.
- Three basic variations of CCD architecture are in common use for imaging systems: full frame, frame transfer, and interline transfer.

REFERENCES

1. A. El Gamal and H. Eltoukhy, "CMOS image sensors," Circuits and Devices Magazine, IEEE, vol. 21, no. 3, pp. 6–20, May 2005.
2. Arai, Toshiki, J. Yonai, T. Hayashida, H. Ohtake and a. T. G. E. Harry van Kuijk. 2013. A 252-V/Lux.s, 16.7 Million Frame per Second 312-kpixel Back Side Illuminated Ultrahigh Speed Charged Coupled Device. IEEE Transaction on Electron Devices. 60(10): 3450–3458.
3. C. Hu, Modern Semiconductor Devices for Integrated Circuits. Prentice Hall, 2010.
4. D. Durini, High Performance Silicon Imaging: Fundamentals and Applications of CMOS and CCD sensors. Elsevier, 2014.
5. D. Poli, "Georeferencing ogf Multi Line CCD Array Optical Sensors with a General Photogrammetric Model," Geoscience and Remote Sensing Symposium, vol. 6, pp. 3908-3910, 2003.
6. D. Poli. 2003. Georeferencing of multi line CCD Array Optical Sensors with a General Photogrammetric Model. Geoscience and Remote Sensing Symposium. 3908–3910.
7. E.E.Michaelides, Particles, Bubbles and Drops; The motion, heat and mass transfer, Singapore: World Scientific, 2007.
8. Etoh, T. Goji, D. Poggeman, G. Kreider, H. Mutoh, A. J. Theuwissen, A. Ruckelshausen and Y. Kondo. 2003. An Image Sensor which Captures 100 Consecutive Frames at 1000000 frames/ second. IEEE Transaction on Electron Device. 50(1): 144–151.
9. Fei, Zhigen, J. Guo, J. Wang and a. G. Luo, "The application of laser and CCD compound measuring method on 3D object detection," 2010.
10. Hong, Zhao, W. Xuan and a. W. Rui, "High speed on line measurement of digital wire outer diameter with laser and CCD technology," Nagoya, Japan, 2003.
11. J. D. Plummer, Silicon VLSI Technology: Fundamentals, Practice, and Modeling. Pearson Education, 2000.
12. J. Goa, Z. Zhang, R. Yao, J. Sun and Y. Zhang. 2011. A Robust Removal Method for Inter Frame Charged Coupled Device Star Images. In Seventh International Conference on Natural Computation.

13. J. Jamaludin, R. A. Rahim, H. B. A. Rahim, N. S. M. Fazli, M. H. F. Rahiman and S. Z. B. M. Muji. 2014. A Review on Optical Tomography System. Jurnal Teknologi. 69(8): 1–6.
14. J. Jamaludin, R. A. Rahim, M. S. Beck, T. Dyakowski, R. A. Williams, F. Rahman, N. A. Nor Muzakkir and M. R. A. Suzzana. 2013. A Review of Tomography System. Jurnal Teknologi UTM. 64(5): 47–51.
15. J. R. Janesick, Scientific charge-coupled devices. SPIE press Bellingham, 2001, vol. 117.
16. J.-Y. Cheng and J.-G. Hwu, "Characterization of edge fringing effect on the C–V responses from depletion to deep depletion of MOS(p) capacitors with ultrathin oxide and high-κ dielectric," IEEE Transactions on Electron Devices, 2012.
17. Li, Q-P., F. Ding and a. P. Fang. 2006. Flash CCD Laser Displacement Sensor. Electronics Letters. 42(16): 910–912.
18. M. Idroas. 2004. Thesis: A Charged Coupled Device Based Optical Tomographic Instrumentation System for Particle Sizing, Sheffield Hallam University.

INDEX

T

U

V

W

3GE Collection on Computer Science: 3D Optical Storage Technology

The best medium for data storage has been a main concern over the past decades. The advent of the information age is accompanied by the rapid growth of digital information, which is in urgent need for the development of huge storage space and security media. Optical storage technology with the huge storage capacity and low cost has become a new choice for information storage. At present, the commercial optical storage devices mainly include Blu-ray discs, digital versatile discs, and compact discs. The main digital circulation has been the optical data storage which has been being improved to accommodate new changes in technology and applications. Optical data storage is an alternative to magnetic disk data storage. Currently data access times are extremely slow for magnetic disks when compared to the speed of execution of CPUs so that any improvement in data access speeds will greatly increase the capabilities of computers, especially with large data and multimedia files. Optical memory is a technology that uses a three dimensional medium to store data and it can access such data a page at a time instead of sequentially, which leads to increases in storage density and access speed. Optical data storage systems are very close to becoming economically feasible. Photo-refractive crystals and photopolymers have been used successfully in experimental optical data storage systems. Such systems exploit the optical properties of these photosensitive materials along with the behavior of laser light when it is used to record an image of an object. 3-D optical data storage technology is one of the modern methods of storing large volumes of data.

This book discusses in details the fundamentals of 3D optical data storage. This includes the features of the 3D optical data storage and the major components that make up the devices. Nonresonant Multi-photon, Sequential multiphoton absorption, microholography and data recording are some of the writing methods used in the 3D optical data storage. The major challenges that are facing these devices as discussed in the book are; media sensitivity, Thermodynamic stability and destructive reading. Two volumetric optical storage approaches, holographic storage and localized-bit storage, were discussed. These approaches provide better capacity, input and output data rates, latency, system volume and low cost in comaprison with the conventional approaches. Storage and retrieval of long data in a relatively smaller space is a challenging task for communication engineer. Now a day's CD's, DVD's, pen derives and hard disk are usually used for this purpose which are not capable holding large amount of data and also retrieval of data takes relatively last time. This book is intended to review the storage of data in 3D optical medium which will hold the large amount of data and will make retrieval easier. The book will help you to study about 3D optical data storage, what is data recording, what is its process, and comparisons with holographic data storage.

ISBN 978-1-98467-702-0

00000

9 781984 677020

BIBLIOTEX
Digital Library

www.3ge-learning.com
email: info@3ge-learning.com