Optoelectronic devices for information storage and processing

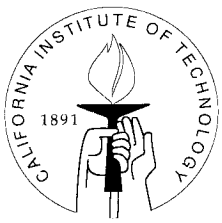
Thesis by

Jean-Jacques P. Drolet

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

1997

(Submitted April 2, 1997)

To Gilles and Monique

Acknowledgements

My gratitude first goes to my adviser, Dr. Demetri Psaltis, for his support, his advice, his patience and his confidence in me throughout my work at Caltech. Dr. Psaltis sparked my interest in optoelectronic devices and volume holography.

One of the early highlights of my research work was a four-month stay in Dr. Jay Patel's laboratory at Bellcore's Navesink Research and Engineering Center in Red Bank, New Jersey. Dr. Patel shared his love of liquid crystals with me. He was always there to answer my numerous questions, and he gave me the freedom to explore and learn as I pleased in his laboratory.

I am indebted to my collaborators George Barbastathis and Ernest Chuang, whose help was invaluable in system-level demonstrations incorporating optoelectronic devices in complete systems.

Dr. Axel Scherer's group at Caltech provided laboratory samples that were employed in my experiments. Dr. Scherer's imagination often seemed limitless; my conversations with him were always stimulating and challenging. I am thankful to him and to his students Samson Timoner, Weihua Xu, Grace Chang, Joyce Wong and Oskar Painter for their help and generosity.

The clean room donated by Gordon and Betty Moore as part of the Moore Laboratory at Caltech was instrumental in fabricating liquid crystal modulators with well-controlled properties.

I had the honor to work with many students and post-doctoral fellows in Dr. Psaltis's group: Chuanyi Ji, Subrata Rakshit, Hsin-Yu Sidney Li, Yong Qiao, David Marx, Robert Denkewalter, Annette Grot, Jiafu Luo, Kevin Curtis, Geoffrey Burr, Allen Pu, Ernest Chuang, Xin An, Michael Levene, George Barbastathis, Chuan Xie, Greg Billock, Greg Steckman, Ali Adibi, Xu Shaw Wang, George Ouyang, Christophe Moser, Elizabeth Ching Ho and Wenhai Liu. I appreciated their friendship and help.

I am grateful to Fai Mok of Holoplex (Pasadena, California), Mark Handschy

and Mike O'Callaghan of Displaytech (Boulder, Colorado), Terry Dorschner and Ed Mangini of Raytheon Electronic Systems (Tewksbury, Massachusetts) and Rahul Sarpeshkar and Thomas Sterling of Caltech for helpful discussions.

My gratitude also goes to YaYun Liu, David Sieving, Lucinda Acosta, Su McKinley, Helen Carrier and Bob Gallagher, who provided the technical, secretarial and administrative help that enabled my projects to make smooth progress.

Support for the research presented in this thesis was provided in part by the Canadian Natural Sciences and Engineering Research Council (NSERC) and the National Science Foundation's Center for Neuromorphic Systems Engineering at Caltech.

Abstract

Optoelectronic information storage and processing systems offer many important advantages compared to their electronic and magnetic counterparts: speed, massive parallelism and insensitivity to interference. Optoelectronic devices are a pivotal technology in the implementation of such systems. Devices consisting of optical inputs and outputs and information processing circuits are needed to interface optoelectronic components and modules to electronic systems, and to perform operations that are more difficult to reliably implement using optics alone. The main thrust of our research is to develop and evaluate optoelectronic technologies conducive to highly integrated optoelectronic components and systems for cost-effective information storage and processing.

At the device level, we describe a simple and inexpensive method for fabricating liquid crystal modulators on silicon integrated circuits. The modulators provide analog amplitude or phase modulation at low voltages. They are compatible with mainstream very-large-scale-integration processes and require only a minimal amount of post-processing performed on conventionally fabricated die. Experimental data are presented and compared to theoretical predictions.

At the chip level, we present an innovative optoelectronic integrated circuit functioning as an optically or electrically addressed spatial light modulator. The device merges the functions of a spatial light modulator and a detector array in a holographic memory system. Moreover, it helps refresh dynamic holograms which slowly decay in a read/write photorefractive memory as a result of their exposure to the reference beam. When combined with the technique of conjugate readout, this device allows a lens-less data path and a very compact, self-aligning integration of the memory module. We also describe two neural arrays, using self-electrooptic-effect devices bonded to a silicon integrated circuit, and light-emitting diodes grown on a commercially processed gallium arsenide integrated circuit.

Finally, at the system level, we describe several integrated system architectures for holographic information storage and processing based on conjugate readout and the aforementioned device. We formulate storage density and cost projections. We report on laboratory prototypes of integrated modular holographic memory. Dynamic holograms were sustained over 50 refresh/decay cycles. Experimental data is presented.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Optoelectronic information storage and processing systems: motivations

1.1.1 Successful optical technologies

The recent years have witnessed the tremendous success of two optical information technologies: optical fibers and compact discs. Optical fibers have replaced copper cables as the medium of choice for high-bandwidth, long-distance communication lines such as telephone trunk lines and wide-area computer networks. The primary reasons for the success of optical fibers are their high speed, low cost and insensitivity to interference. They are also gradually replacing copper connections in local area networks (LAN) because of their decreasing cost and of the increasing bandwidth demands of modern computer applications and the proliferation of multimedia communications (e.g., live video teleconferencing). Fibers are also spreading deeper into the structure of telephone networks. Analysts forecast that fiber-to-the-curb and fiber-to-the-home will be commonplace in a matter of a few years.

Shortly after its introduction in 1984, the compact disc (CD) had completely supplanted vinyl records as the medium of choice for distributing music. A few years later, the compact-disc read-only memory (CD-ROM), which employs the same physical technology and format, complemented by a set of file system standards, became the primary distribution medium for computer programs and databases. Compact discs record information in the form of tiny pits in a polycarbonate substrate. The pits are detected by a focused infrared laser beam. CDs are small (120 mm in diameter and 1.2 mm in thickness) and lightweight because the optical technology upon which they are based allows dense data recording (the minimum pit length is 0.834 μm

and the track pitch is 1.6 μm). One of their main advantages over vinyl records is that CDs record digitized sounds and employ interleaving and error-control schemes, resulting in high fidelity and no gradual degradation of the recorded information. Only optical storage could economically provide the density needed for large storage capacity, error-control overhead and fast access time in such a small package. It is estimated that the fabrication of a CD costs less than $1 a piece. This low cost has been a key factor in the technology's rapid and deep penetration of the music and computer media markets. Compact discs are also very rugged because the information is recorded well under their surface, in contrast to vinyl records and magnetic discs and tapes. As of this writing, an improved compact disc standard, called *Digital Versatile Disc* or DVD, is being readied for the limelight. The new standard calls for two-sided, two-layered discs that store up to 17 GB of information. Its proponents target, among others, the video-rental and -retail market. The volume fabrication cost of DVDs is expected to be as low as 36% of that of a VHS tape[1].

1.1.2 Parallelism

The success of optical fibers and compact discs has been enabled by properties of optical systems, and developments in the fields of devices and manufacturing techniques. Fibers take advantage of the high speeds and robustness to interference provided by lightwave communication, while CDs tap on the density made possible by optical data storage and gain robustness from the propagation of the readout beam in free space and in the polycarbonate substrate, which allows the protection of the pits encoding the data by the substrate. Both technologies have greatly benefited from the development of laser diodes and from inexpensive fabrication and assembly techniques (in the case of fibers, reliable and simple splicing techniques; in the case of the compact disc, stamping and low-cost modular integration of the optical system in the drive).

Despite the success of these technologies, one important property of optical systems remains mostly untapped in either fiber communication systems or compact discs: the massive *parallelism* enabled by optics. Optical fibers transmit data in a

bit-serial fashion[1]; likewise, data is read from CDs one bit at a time. Optical systems can process and communicate[2] two-dimensional data pages (e.g., images) in parallel. However, optics excels at complex interconnection schemes (using, e.g., holograms) and linear operations such as weighted sums and Fourier transforms[2]. One can therefore envision the widespread use of powerful and rich computation and storage primitives such as parallel-access holographic memories, optical correlators for pattern recognition, programmable optical interconnection systems and optical neuromorphic systems (artificial neural networks), processing 2-D information frames in parallel, and taking advantage of the other properties of optics (speed, robustness to interference, propagation in free space, dense information storage, etc.). We will exemplify the potential power of optical parallelism using holographic data storage, optoelectronic correlators and neuromorphic systems. These powerful optical paradigms based on volume holography constitute basic building blocks of high-performance parallel information storage and processing systems. Then, in Sections 1.2 and 1.3, we expose some of the challenges that must be faced in order to enable such systems to evolve into competitive commercial realities.

1.1.3 Holographic data storage

One way to take advantage of the parallelism offered by optics in information storage and processing systems is to employ holographic data storage. Figure 1.1 shows a simple holographic storage system. Holograms are stored in the recording medium by the interference of a *signal beam* (coming from the upper left corner of the figure) with a *reference beam* (from the upper right corner). Data are imprinted on the signal beam by a spatial light modulator (SLM). Many holograms can be superimposed in the same volume by slightly rotating the direction of propagation of the reference beam

[1] Many communication paths can share a single fiber (through, e.g., time-division or wavelength-division multiplexing), because the bandwidth of the fiber is typically much higher than that required by a single channel (e.g., a telephone conversation). Although this multiplexing can be seen as a form of parallelism, we are mainly concerned here with the 2-D spatial parallelism offered by optics, which stems from a 3-D optical communication channel's ability to process 2-D information sets (e.g., images) in parallel.

[2] Through free space, lenses and other optical elements.

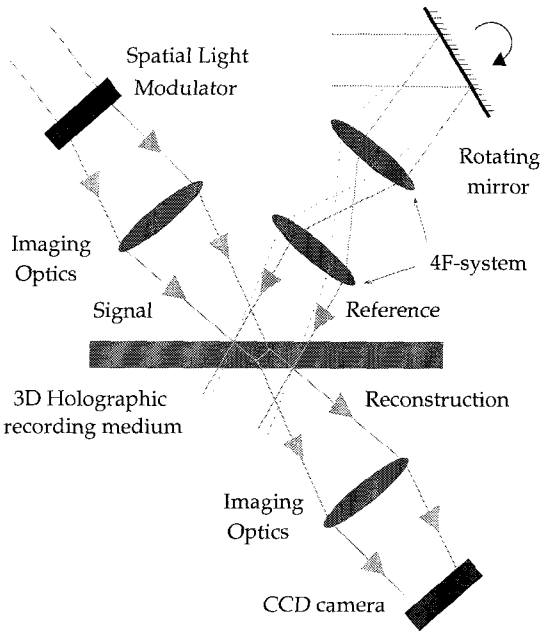


Figure 1.1: A simple holographic storage system.

between recordings[3]. In Figure 1.1, this is accomplished using a motorized rotating mirror. Detuning the reference beam causes a Bragg mismatch in the reconstruction; a new hologram can be recorded at a null of the Bragg selectivity curve (diffraction efficiency vs. reference propagation angle). In the transmission geometry illustrated in Figure 1.1, the minimum angular spacing between holograms is given by[3]

$$\Delta\theta = \frac{\lambda}{2L\sin\theta}, \tag{1.1}$$

where λ is the wavelength of the beams and L is the thickness of the recording medium. The thicker the medium, the stronger its angular selectivity. In the 90° geometry, in which the propagation axes of the signal and reference beams are normal to each other, $\Delta\theta \approx \frac{\lambda}{L}$.

The physical nature of the recording process can take many different forms, depending on the material being used. In photorefractive crystals[4, 5], electrons are

[3]Angular multiplexing is but one of the many multiplexing techniques. Several more will be discussed in Chapter 5.

excited to the conduction band in bright areas of the fringes of the interference pattern between the signal and reference beams. The electrons diffuse toward the dark areas, where they are trapped by impurity atoms (e.g., Fe atoms in $LiNbO_3$). The electrons remain trapped long after the beams have been removed. The resulting electric field gives rise to a non-uniform refractive index through the electrooptic effect. In addition to diffusion, drift and the photovoltaic effect can also participate in the photorefractive effect, as they do in $LiNbO_3$:Fe. Holograms recorded in photorefractive crystals can be *fixed* (e.g., made non-volatile) thermally[6, 7] or electrically[8], depending on the nature of the recording material.

A different mechanism is responsible for recording non-volatile phase holograms in photopolymers such as DuPont's HRF-150[9]. In this case, exposure to light catalyzes a chemical reaction in the material. The fringe pattern resulting from the interference between the signal and reference beams again gives rise to a phase grating.

During readout, the mirror is rotated to match its position during recording of the desired data page, and the reference beam is shone alone. The signal beam is *reconstructed* by the diffraction of the reference beam by the grating recorded in the holographic material. It propagates as if it emanated from the SLM, toward a detector array such as a CCD camera. Note that both the recording and readout operations are performed in parallel, i.e., many bits at a time. For example, with a 1000×1000 SLM, one million bits can be written *in parallel.* This operation can be repeated at each of the angular "locations" defined by the positions of the rotating mirror in the reference beam path. Readout is also performed in parallel. The storage density theoretically achievable using volume holography is of the order of V/λ^3, where V is the volume of the recording medium and λ is the wavelength of the light beams[10]; however, in practical systems, the density is much lower because of geometrical restrictions (not all possible gratings can be written in any given configuration) and dynamic range limitations. Chapter 5 describes several compact, modular and integrated holographic memory architectures and provides storage density and cost estimates.

1.1.4 Correlators

In the holographic memory system shown in Figure 1.1, the signal beam corresponding to a particular data page is reconstructed by shining the reference wave, propagating at the appropriate angle, alone. What happens if, instead, the signal beam is shone alone, with the SLM displaying an arbitrary input pattern? Intuitively, one could think that the result would be some sort of reconstruction of the reference beams employed to record the various holograms (stored images, or templates). More accurately, these "reconstructions," once focused[4] by a lens, are found to be the correlation functions between the input pattern and each of the templates[5]. Thus, a holographic memory system can be transformed into a *correlator*, a device commonly employed in pattern recognition tasks, with only a few minor modifications.

Figure 1.2 shows a volume holographic correlator, which is straightforwardly derived from the holographic memory system illustrated in Figure 1.1. The correlation peaks, which can be thought of as a measurement of the degree of similarity between the input and each of the stored templates, appear in the output plane. It is assumed that the templates have already been stored as non-volatile holograms in the holographic medium. For example, they could have been recorded in photopolymer, or recorded and fixed in a photorefractive crystal. Alternatively, the templates could be recorded as dynamic holograms; in this case, the set of templates could be updated during the operation of the system, in such a way as to allow the system to adapt to different input conditions. Chapters 3 and 5 (§5.4) describe a device and an architecture that allow a compact and elegant system to exhibit this adaption property while preventing an excessive decay of the previously stored templates.

Holographic correlators were first proposed and demonstrated in the 1960's[11]. The early experiments employed thin holographic media in which a single template was stored. Holographic correlators exhibit the same kind of 2-D parallelism as holographic memories: the entire correlation function is computed in parallel, over the extent of

[4]Or, more accurately, Fourier-transformed.

[5]Provided that the SLM is in the front focal plane of the lens in the signal path, and that the recording medium is in its back focal plane.

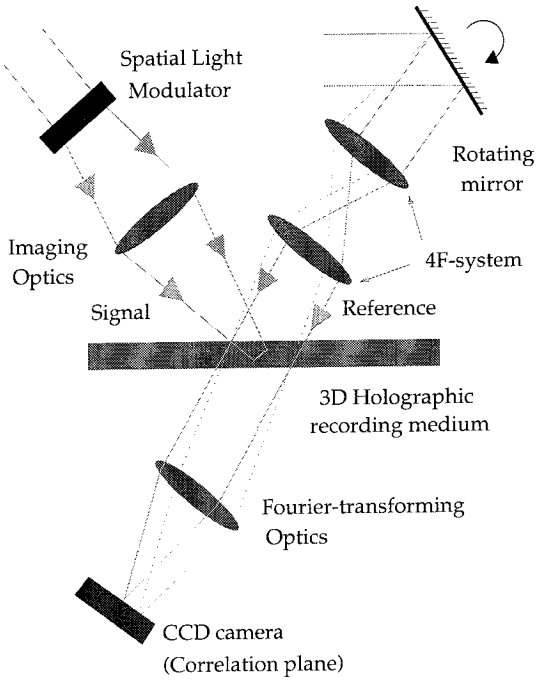


Figure 1.2: A volume holographic correlator.

the input image and the template. In comparison, on a conventional digital computer, computing a 1-D correlation with a known template requires two Fourier transforms and the scalar multiplications of two image matrices, or $\mathcal{O}(N \log_2 N)$ operations[12], where N is the number of pixels.

The use of thick media endows the holographic correlator with a second level of parallelism. Many templates can be stored in the recording medium, and all the corresponding correlation functions are computed simultaneously. In this case, the price to pay is a loss of shift invariance. While the output of a correlator employing a thin medium is a true correlation function[6], possessing full shift invariance in two dimensions[13], the shift invariance along the multiplexing direction (the direction of strong Bragg selectivity) in the volume holographic correlator is limited because the correlation peaks taper off, following the selectivity curve, when the input is moved. The trade-off between template storage capacity and shift invariance is determined by the thickness of the recording material and by the geometry of the system (see

[6]Within the validity limits of Fourier Optics[2].

equation 1.1). It is also possible to control the trade-off by slightly removing the center of the holographic recording medium from the back focal plane of the input lens[14][15, Appendix].

1.1.5 Optoelectronic neuromorphic systems

The elegance and simplicity of optical correlators naturally leads to optoelectronic artificial neural networks[16]. Such information processing systems, which are loosely patterned after biological neural structures, consist of a large number of densely interconnected, simple processors[17]. The behavior of an artificial neural network does not need to be fully specified *a priori*; the strength of the interconnections (called *interconnection weights*) between its processing units is "learned" by presenting a sequence of training inputs and desired outputs. Such networks exhibit interesting generalization properties. They have been applied to various complex tasks such as handwriting recognition[18], stock market prediction[19] and credit-worthiness evaluation.

For example, Figure 1.3a shows a two-layer[7] neural network. An input pattern is presented to the network as the vector $\mathbf{f}$. $\mathbf{f}$ could represent, e.g., the intensities of the pixels in an image to be recognized. The first layer of interconnections links input units to each of the units in the *hidden* layer. The input to a unit i in the hidden layer is a weighted sum of components of the input pattern,

$$\sum_j w_{ij}^{(1)} f_j, \tag{1.2}$$

where $w_{ij}^{(1)}$ is the strength of the interconnection between input j and neuron i in the hidden layer. The neuron applies a nonlinear function $g^{(1)}(x)$ to the weighted sum; a commonly used function is the sigmoid $g^{(1)}(x) = \tanh x$. The output of the unit is

$$h_i = g^{(1)}\left(\sum_j w_{ij}^{(1)} f_j\right). \tag{1.3}$$

[7]It is customary to count the layers of interconnections, or the layers of non-input units.

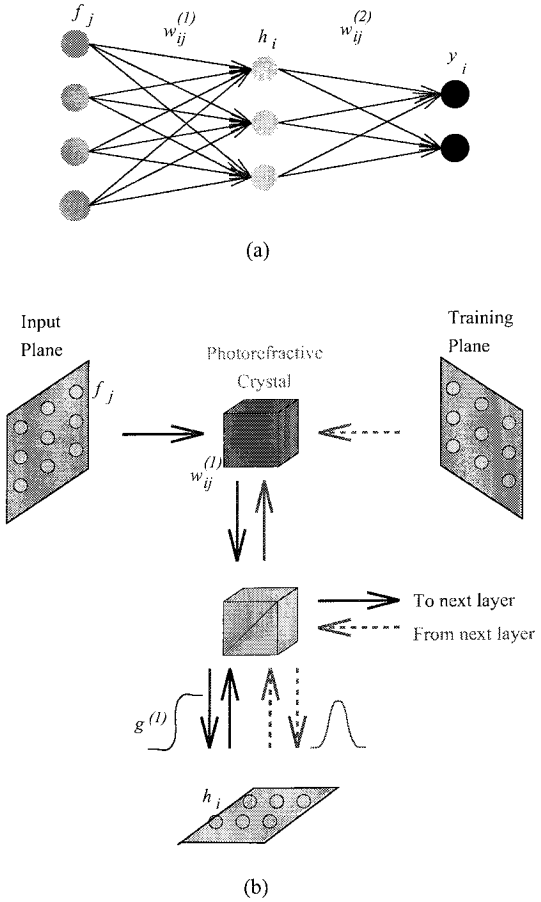


Figure 1.3: Two-layer artificial neural network (a) and optoelectronic implementation of one layer of interconnections (b).

The second layer of interconnections links units in the hidden and output layers. The output of the network is a vector $\mathbf{y}$, where

$$y_k = g^{(2)}\left(\sum_i w_{ki}^{(2)} h_i\right). \tag{1.4}$$

The role of the hidden layer is to provide a linearly separable representation of the input space, so that a single additional layer can provide any desired classification. Networks such as the one shown in Figure 1.3a can be trained using the famous back-propagation algorithm[17].

Optics offers many attractive properties for the efficient implementation of artificial neural networks such as the one illustrated in Figure 1.3a. Optical interconnections naturally operate in parallel at the speed of light, and are immune to RF interference and capacitive coupling. Another valuable advantage of optical interconnections is that they provide a solution to the wiring bottleneck common in densely interconnected VLSI circuits. Because VLSI die are primarily two-dimensional structures, densely interconnected circuits (such as, e.g., complex programmable logic devices (CPLDs) and artificial neural networks) tend to devote most of their area to interconnection wires. Optical interconnections relieve this problem by allowing 2-D arrays of signals to propagate in 3-D free space.

Weighted sums (equation 1.2) are also naturally implemented using optics. For example, one of the earliest optical implementations of a neural network[20] employed a simple vector-matrix multiplier to compute the desired weighted sums. The fixed interconnection weights were recorded on a mask. The nonlinear neuron response function is most often implemented using an optoelectronic device such as an optically addressed spatial light modulator or an optoelectronic integrated circuit instead of purely optical components. Optoelectronic devices generally offer more flexibility and make it easier to implement the desired function in an inexpensive, well-controlled and power-efficient fashion. We will return to this topic in Section 1.2.

Photorefractive crystals are a popular medium to store interconnection weights and topologies[21]. They provide a vast storage capacity. As mentioned previously,

a photorefractive crystal of volume V can theoretically store up to $\sim V/\lambda^3$ independent optical interconnections at a wavelength λ; however, geometric and dynamic range limitations reduce this number by several orders of magnitude in practice. The storage of 160,000 holograms at $\lambda = 488$ nm in $\sim$ 1 cm^3 of $LiNbO_3$ was recently demonstrated[22]; if each hologram contains one million pixels (or interconnection gratings), this system stores 1.6×10^{11} independent interconnections. Interconnection weights stored in photorefractive crystals can also be updated *in situ*, in real time, making *adaptive* optical neural networks[23] a possibility.

The correlator shown in Figure 1.2, with a few modifications, forms the basis for an optoelectronic neural network layer. This is because the center of a correlation function (overlap integral) is simply the inner product given by equation 1.2. Figure 1.3b schematically shows a modular optoelectronic implementation of one layer of a multilayer neural network such as the one in Figure 1.3a. The Fourier-transforming lenses (see Figure 1.2) were omitted for the sake of simplicity. Input patterns are presented to the network on a spatial light modulator in the input plane. The interconnection topology and weights are implemented using the photorefractive crystal in the 90° geometry. The resulting weighted sums are projected on neural array at the bottom of the figure. This device could consist of an optoelectronic integrated circuit or an optically addressed spatial light modulator, for example. It applies the desired nonlinear function (e.g., a sigmoid) to the weighted sums and passes its output on to the next layer.

The same system configuration is also employed during learning[23], with the addition of a training plane. In this case, the training plane input interferes with the back-propagated[17] error signal to update the stored weights. Because the error signal propagates backward, the input must also be presented "backward" to ensure consistency with the interconnection gratings employed during forward propagation, hence the need for a separate training plane. This plane is not required when learning is complete.

An optoelectronic integrated circuit implementing all the functionality required of hidden or output units was demonstrated[24]. It used ferroelectric liquid crystal

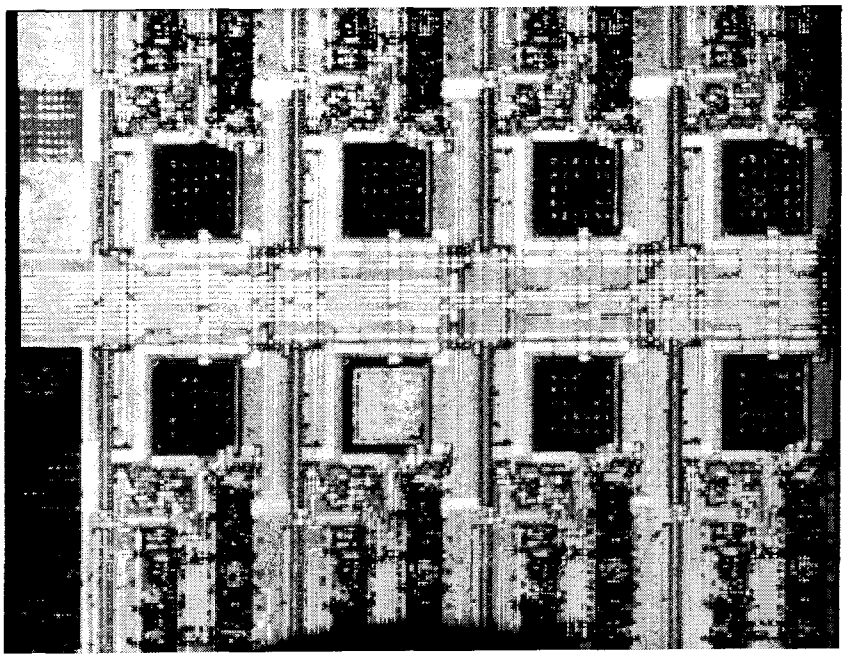


Figure 1.4: Optoelectronic neural array with FLC-VLSI modulators.

modulators assembled on a VLSI die fabricated in a commercial 2.0 μm CMOS process through MOSIS. Figure 1.4 shows the eight-neuron device, with one of the modulators turned on.

1.2 Optical and optoelectronic components

Section 1.1.2 made the case that the potential for massive parallelism in optical systems remains largely untapped in high-volume commercial optical and optoelectronic systems. This Section and the next will enumerate and explain the opportunities that can be exploited and the challenges that must be faced in order to enable the development of high-performance optoelectronic systems with parallel data processing, storage or retrieval. Section 1.3 will describe architectural considerations leading to compact, high-performance optoelectronic systems. However, let us first discuss opportunities and requirements related to optical and optoelectronic components.

1.2.1 High-density integration and smart pixels

Optical information storage and processing systems stand to benefit from highly integrated optoelectronic arrays. Such devices can interface an optical system to a conventional electronic computer, or implement functions that would be difficult to realize using optics alone. For example, in Section 1.1.5, we saw that optics' ability to efficiently compute linear operations such as inner products leads to elegant and natural optical implementations of artificial neural networks; however, non-linear local operations such as neural responses were best implemented using an optoelectronic device, such as the one shown at the bottom of Figure 1.3. Examples of interfacing devices are electrically addressed spatial light modulators and detector arrays, which perform electronic-to-optical and optical-to-electronic transduction, respectively. Large-scale integration allows a large number of pixels to be processed in parallel by a single optoelectronic component. Chapter 3 will describe an optoelectronic device that performs simple local processing and two-way optical-electronic transduction in holographic systems.

In many cases, optoelectronic arrays are based on very-large-scale-integration (VLSI) fabrication techniques employed on silicon or gallium arsenide (GaAs) wafers. Such optoelectronic arrays are called optoelectronic integrated circuits (OEICs). Both silicon and gallium arsenide are heavily employed by the semiconductor industry. Using them and their associated processing technology allows optoelectronic devices to leverage the high-volume advanced processes developed by the semiconductor industry, and hence their associated high device densities, high yields, low costs, and large usable device areas. To a large extent, such optoelectronic devices can ride the coattails of regular electronic integrated circuit development, e.g., Moore's law[8]. The reliability and small feature sizes of commercial VLSI processes enable the integration

[8]In some cases, however, electronic advances can impair the suitability of a process for the fabrication of OEICs. One notable example is the silicidation of polysilicon and drain/source diffusions, designed to reduce resistance in advanced CMOS processes, which can make the fabrication of sensitive photodetectors nearly impossible. Fortunately, many manufacturers add silicide blocking masks to their process. It is interesting to note that the addition of these masks was usually not motivated by optoelectronic requirements, but rather by the need for a means to fabricate high-quality compact resistors in mixed-mode integrated circuits.

of circuitry within the elements of the array (referred to as pixels) or in the periphery of the array. Intra-pixel integration allows each pixel to perform local information processing or storage tasks; the array elements are then dubbed *smart pixels*. The addition of circuitry in the periphery of the pixel array can be employed to monolithically integrate power, control and signal processing functions with the optically active part of the device, resulting in a compact, highly integrated and potentially faster and cheaper device. While the bare device fabricated in a commercial process may not possess all desired optical input/output functions, such functions can often be added using a minimum of post-processing. Examples will be given in Chapters 2 and 4.

Silicon and gallium arsenide are attractive to OEIC designers for different reasons. The silicon market far outweighs that of gallium arsenide. Silicon is easier to grow with good quality. The surface of silicon can be oxidized to form silicon dioxide, an excellent insulator with a large dielectric breakdown field. Very thin layers of silicon dioxide (typically 1 to 50 nm thick) are employed as the insulator between the gate and the substrate in MOS transistors. Silicon lends itself well to the fabrication of very dense, relatively inexpensive, low-power integrated circuits. The main motivations for using silicon in optoelectronic integrated circuits are cost, density and compatibility with the ever-improving silicon processing technology employed to fabricate purely electronic integrated circuits. Silicon has a bandgap of 1.12 eV at room temperature, and is optically sensitive throughout the visible range and the near infrared. Nevertheless, silicon has one major drawback as a general purpose OEIC material: it is difficult to provide a silicon device fabricated in a conventional process with optical outputs without performing some post-processing. Silicon cannot emit light with any efficiency because of its indirect bandgap, which would require a lattice interaction accompanying each band-to-band recombination in order to satisfy momentum conservation. The material also does not lend itself well to the fabrication of multiple-quantum-well structures due to lattice mismatches.

Gallium arsenide possesses two properties that make it a particularly interesting material for optoelectronic applications. First, its direct bandgap enables the efficient emission of light. Second, other elements of columns III and V of the peri-

odic table (Al and In from column III; P from column V) can be substituted for a fraction of the Ga or As atoms in GaAs, without introducing a significant lattice strain. This property enables the fabrication of efficient laser diodes and of fast light modulators; it is due to the similarity of the lattice constants of GaAs and of related compounds, such as AlAs. It is possible to grow superimposed thin layers of GaAs and related compounds (such as GaAlAs) using techniques such as liquid phase epitaxy (LPE), metallo-organic chemical vapor deposition (MOCVD) and molecular beam epitaxy (MBE). The latter can achieve extremely fine thickness, down to a few atoms. The substituted atoms can be employed to control the bandgap and the index of refraction of the compound. In double heterostructure laser diodes, for example, a thin layer of low-bandgap, high-index material is embedded in a higher-bandgap, lower-index material[25, Chapter 15, for example]. One common structure is a GaAs active layer sandwiched between layers of GaAlAs. The bandgap difference confines injected charge carriers in the active region, where they recombine to produce light. The light is in turn confined by the index step. Another example is *multiple-quantum-well (MQW) modulators*, optoelectronic devices based on bandgap modulation that provide an electrically controlled optical absorption. Optoelectronic integrated circuits based on gallium arsenide leverage investments made by the semiconductor industry to develop the fabrication of light-emitting diodes, laser diodes and gallium arsenide integrated circuits. The semiconductor industry's main interest in gallium arsenide technology is speed. The mobility of electrons at room temperature in pure gallium arsenide is 8000 $cm^2/V \cdot s$, compared to 1400 $cm^2/V \cdot s$ in pure silicon. However, processing difficulties, along with the lack of a thermal oxide or some other high-quality dielectric for metal-insulator-semiconductor (MIS) structures, have prevented the widespread acceptance of gallium arsenide integrated circuits in all but the most demanding high-speed applications (e.g., gigahertz communication transmitters). As a result, gallium arsenide devices tend to be more expensive than their silicon counterparts. The most commonly employed device in gallium arsenide integrated circuits is the metal-semiconductor field effect transistor[26, 27] (MESFET), a four-terminal device similar to a junction field-effect transistor (JFET), with a rectify-

ing metal-semiconductor contact substituted for the p-n junction. Unlike MOSFETs, MESFETs exhibit a non-zero quiescent current. Moreover, many processes offer only n-channel MESFETs, because, unlike electrons, holes do not possess a particularly high mobility in gallium arsenide[9]; process development efforts are focused on the device type that offers the most important benefits compared to competing technologies. Together, these properties make it more challenging to design low-power analog or mixed-mode integrated circuits using gallium arsenide instead of silicon. Nevertheless, significant developments have been made, notably in the field of gallium arsenide neural arrays[29, 30, 31, 32].

In summary, silicon OEICs take advantage of the rich features, low static power dissipation and low cost offered by mainstream silicon VLSI processes, whereas gallium arsenide offers efficient light emitters (including lasers), higher speed, and a means to modulate its index of refraction and its bandgap. In many cases, one of the two materials will clearly stand out as being best suited to the application at hand. However, it is also possible to benefit from the best features of both materials by designing devices that use both silicon and gallium arsenide. One of the best examples of this approach is Lucent Technologies' Hybrid-SEED process[33], which combines mainstream sub-micrometer silicon electronics with gallium arsenide multiple-quantum-well structures that can be employed as fast modulators or as high-efficiency, fast p-i-n photodiodes. Flip-chip bonding is employed to attach a gallium arsenide die containing MQW devices to a conventionally processed silicon die containing most of the electronic components. The gallium arsenide wafer is then removed, leaving only the MQW diodes bonded to the silicon die. The MQW diodes connect to the silicon circuits via solder bumps contacting the topmost level of metal on the silicon die. It is possible to integrate circuits using all the normal features of the silicon process except the topmost metallization level under the MQW devices. Smart pixel arrays with good pixel density and fill factor can be fabricated using this process. We will describe a Hybrid-SEED neural array in Chapter 4.

[9]The mobility of holes in pure gallium arsenide at room temperature, 390 $cm^2/V \cdot s$, is lower than that in pure silicon, 480 $cm^2/V \cdot s$[28].

Optoelectronic devices can be classified according to the functions they perform: optical detection, optical output and information processing and storage. One way to take advantage of the high levels of integration enabled by modern VLSI processes is to combine two of more of these functions. For example, a pixellated optically addressed spatial light modulator contains optical detectors and modulators within each of its pixels. Most smart pixels are comprised of optical inputs, optical outputs and some local information processing or storage circuitry. Integration can also be beneficial to devices performing only optical detection or only optical output. For example, high-level multiplexing can be achieved in high-resolution, high-quality liquid crystal displays by integrating a transistor in each pixel of a display device, as is done in thin-film-transistor (TFT) active-matrix displays. The gating transistor frees the display unit of the multiplexing limits associated with passive-matrix displays[34, 35]. The integration of additional transistors can also confer more flexibility to light detection devices. For example, in simple charge-coupled devices, the entire pixel array is scanned during each frame period. Pixel information can only be read sequentially. However, random-access detectors can be built by adding gating transistors within the pixels of a photodetector array, and addressing circuitry in the periphery of the array. Pixel values can then be read in any desired order. The optical signals can be integrated for different amounts of time in different regions of an image, a desirable feature when information must be extracted with high fidelity from both bright and dark areas. Adding part of a simple source-follower amplifier within each pixel[36, 37] allows faster low-noise operation[10] and non-destructive readout of the optical signals[38], in addition to random addressing. Highly integrated sensor arrays include on-chip analog-to-digital converters[39], and can integrate an entire camera[11] on a chip[40].

[10]Because the column buses are driven by an amplifier rather than the photodetector itself.

[11]Without the lens, of course.

1.2.2 Optical output devices

Optical output devices can be classified in two categories: emitters and modulators. The former category is comprised of devices, such as light-emitting diodes and laser diodes, that generate their own optical beams, while the latter consists of elements that modulate a property (phase or amplitude) of an externally generated beam. A smart pixel array using emitters (LEDs) will be described in Chapter 4. Devices based on emitters can lead to more self-contained systems, since they do not require an external optical power source. However, particular attention must be paid to power dissipation issues. Moreover, in holographic systems, which are the focus of most of this thesis, coherence is an issue. While each laser diode in an array may possess an excellent temporal coherence, in general, the array as a whole is not guaranteed to be spatially coherent. In order to write holograms with a data beam emanating from such an array, the array must be locked into coherence, e.g., using a feedback scheme.

Because of our focus on holography, we will concentrate our attention on modulators. Optical modulator technologies, especially those suitable for integration with semiconductor die, are an effervescent field of study. A number of different technologies are available, offering different trade-offs between speed, contrast, power consumption, and operating wavelength.

1.2.3 Materials

In Section 1.1.2, we introduced holographic data storage, holographic correlators and optoelectronic holographic neuromorphic systems as building blocks of high-performance optoelectronic information storage and processing systems. Volume holography is a key technology in all these areas. One of the main challenges in unleashing its full potential is the development of recording materials with high sensitivity, diffraction efficiency, optical quality and storage time, and low losses. The materials should also be optimized for operation in a spectral range where efficient, low-cost sources are available.

To date, the most popular holographic recording materials are inorganic photore-

fractive crystals, such as $LiNbO_3$, $BaTiO_3$, $Sr_{1-x}Ba_xNb_2O_6$ (SBN), $KNbO_3$ and GaAs. The photorefractive effect[4, 5] is based on charge generation, transport and trapping, and on the electrooptic effect. When two coherent light beams interfere in a photorefractive material, charges are generated in the bright areas of the interference pattern. The charge carriers migrate toward the dark areas, where they are trapped, leaving behind uncompensated ions (see Figure 1.5). The non-uniform charge distribution gives rise to a space-charge field, with a 90° phase shift with respect to the charge distribution, which eventually stabilizes the process by causing a local drift current that opposes the optically induced charge migration. The space-charge field in turn gives rises to a refractive index grating through the electrooptic effect. The net result is a replica of the interference pattern, in the form of an index grating in the recording medium. The holograms can be erased by exposure to uniform light. They also decay in dark storage conditions, with time constants varying from seconds to months, depending on the temperature, the material, the dopant species, and the doping level. This volatility of photorefractive holograms is a mixed blessing. On the one hand, it allows holograms to be erased and rewritten, or to be continuously updated; this is useful is read/write holographic memories and adaptive neural networks. On the other hand, stable holograms are desirable in most applications, except in update cycles (e.g., rewriting a page in a memory system, or updating the interconnection weights associated with a neuron in a neural network). Moreover, light-induced decay limits the maximum diffraction efficiency that can be obtained when several holograms are multiplexed in a common volume. This latter point deserves a more detailed explanation.

When multiplexed holograms are recorded in succession in a common volume, the previously recorded holograms decay during each recording operation. The diffraction efficiencies of the holograms must be equalized in order to ensure that they can all be detected with high fidelity and by the same sensor. This is accomplished by employing an *exposure schedule*. Intuitively, we expect that the first holograms will require a longer exposure to compensate for their prolonged erasure. The strength[12] w_m of the

[12]The amplitude of the photorefractive space charge field.

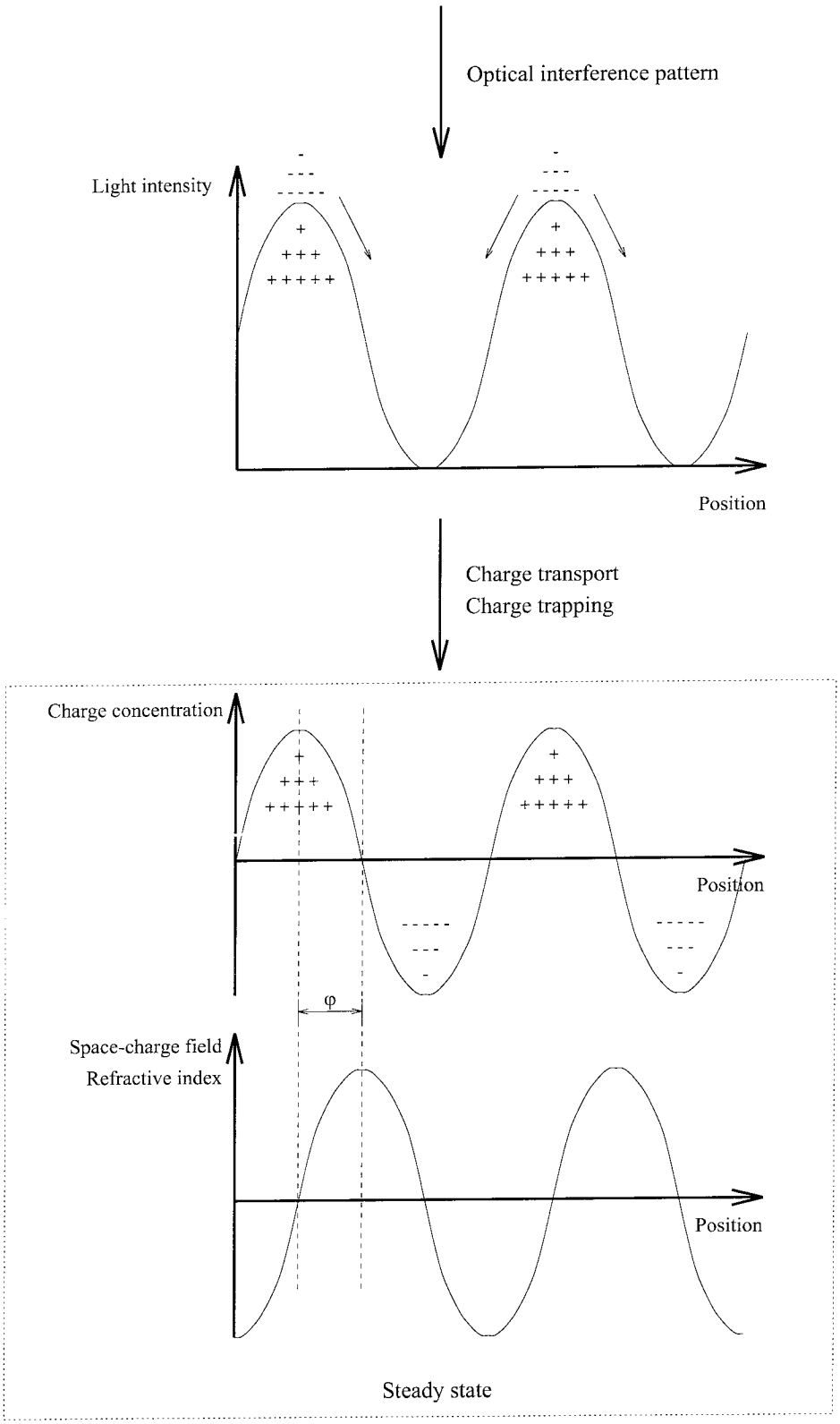


Figure 1.5: Mechanism of the photorefractive effect.

mth hologram, after a set of M holograms have been recorded, can be expressed as

$$w_m = A_0 \left(1 - e^{-\frac{t_m}{\tau_w}}\right) e^{-\sum_{m'=m+1}^{M} \frac{t_{m'}}{\tau_e}}, \tag{1.5}$$

where A_0 is the maximum strength of a single hologram, t_n is the recording time of the nth hologram, and τ_w and τ_e are respectively the recording and erasure time constants[13]. The condition that all the holograms possess the same strength can be expressed as follows:

$$\forall m \in \{1, 2, 3, \ldots, M-1\}, w_m = w_{m+1}.$$

The solution, assuming that the recording times are much smaller than the recording and erasure time constants, is[14]

$$t_m = \frac{\tau_e}{m},$$

and the equalized hologram strength is

$$w \approx \frac{A_0 \tau_e}{M \tau_w}.$$

The diffraction efficiency is proportional to w^2:

$$\eta \propto \left(\frac{A_0 \tau_e}{M \tau_w}\right)^2, \tag{1.6}$$

or, defining[41] $M/\#$ (pronounced *M-number*) as

$$M/\# \equiv \frac{A_0 \tau_e}{\tau_w}, \tag{1.7}$$

[13]In some materials, including $LiNbO_3$:Fe, the two time constants are different.

[14]When this simple exposure schedule is used, the first few holograms may not be properly equalized because their recording times are not much smaller than the time constants. Alternative exposure schedules, possibly using pre-computed look-up tables, can make all holograms equally strong.

$$\eta \propto \left(\frac{M/\#}{M}\right)^2. \tag{1.8}$$

We see that the diffraction efficiency of individual holograms goes as the inverse of the square of the number of holograms. The result also applies to *incremental recording*, where small exposures are successively applied to individual holograms over a large number of cycles.

$M/\#$ is a lumped figure of merit of a holographic system. It depends primarily on the recording material, but also on other system parameters (geometrical factors, relative strengths of the beams, ...). If absorption is neglected, $M/\#$ is proportional to the interaction length between the beams. Therefore, it makes sense to define an $M/\#$ per unit length:

$$(M/\#)^* \equiv \frac{\partial}{\partial z}(M/\#), \tag{1.9}$$

where z is the coordinate along the interaction axis.

Photorefractive crystals can be grown to relatively large sizes (a few cubic centimeters) with good optical quality. $LiNbO_3$:Fe is currently the most popular in holographic storage applications. It is sensitive at the short-wavelength end of the visible spectrum, and is commonly employed at the output wavelengths of argon ion lasers (488 nm, 514 nm). Experimental setups operating at 488 nm in our laboratories typically exhibit $M/\# \sim 1.5$. However, $M/\# \sim 11$ can be achieved[42]. Dopants other that Fe (Ce and Rh, in particular) are currently being investigated as means of improving the sensitivity and other properties of $LiNbO_3$. $M/\#$ decreases with increasing wavelength.

Holograms written in $LiNbO_3$:Fe can be made non-volatile by exposure to a moderately elevated temperature. This process, which involves the copying of the electronic grating onto an ionic grating, is known as *thermal fixing*[6]. Fixed holograms can be read without erasure.

Another relatively popular material is $BaTiO_3$, which possesses high sensitivity throughout the visible spectrum, and thus records holograms fast. Unfortunately, $BaTiO_3$ typically also possesses a low τ_e (erasure time constant in the presence of

recording beams) and a short dark decay time constant (a few seconds). However, recent measurements indicate that cerium-doped $BaTiO_3$ ($BaTiO_3$:Ce) could exhibit a dark decay time as long as 2200 years at room temperature[43]. The response time of $BaTiO_3$:Ce with 15 ppm cerium has been measured as 220 ms at an intensity of 800 mW/cm^2 at 514.5 nm. Unfortunately, crystal growth difficulties usually limit the useful size of $BaTiO_3$ crystals to 0.1 cm^3 to 1 cm^3.

SBN, which is a ferroelectric material, also possesses high sensitivity. Photorefractive holograms written in SBN can be electrically fixed[8]. The process involves the flipping of ferroelectric domains under the influence of an externally applied electric field, locally enhanced by the photorefractive space-charge field. The main difficulty associated with SBN is obtaining samples with sufficient optical quality. SBN crystals typically exhibit significant striations.

We briefly discussed above the thermal and electrical fixing processes that transform dynamic or transitory electrical gratings into permanent ionic grating or ferroelectric domain gratings. Three other methods of non-volatile storage are also applicable to photorefractive crystals. The first method employs different wavelengths for recording and readout[44]. Holograms are written using a wavelength at which the recording material is sensitive, and read out at a wavelength to which it is insensitive. The different wavelengths cause a spatially non-uniform Bragg matching, whose detrimental effects can be overcome using a suitable interleaved page formatting technique.

The second method, called *gated recording*, also involves two wavelengths[45]. The signal and reference beams employ the same wavelength during both recording and readout operations, thereby eliminating the non-uniform Bragg matching mentioned above. However, a third beam at a different wavelength is shone simultaneously with the signal and reference beams during recording operations[15], and causes a transient sensitization of the recording medium to the wavelength of the signal and data beams. The material is otherwise insensitive to this wavelength; thus, readout is non-destructive.

[15]The third beam can also be shone shortly before the recording operation.

A third solution, applicable to virtually any recording medium with dynamic holograms, is periodic refreshment. Each decaying hologram is periodically sensed and rewritten[46]. We will outline the refreshment process in §1.3 and show that it can be performed without requiring additional components in a compact and elegant holographic memory module.

The recent years have also witnessed significant developments in the field of organic holographic recording materials. In principle, their properties can be adjusted to the requirements of specific applications with relative ease by slightly modifying parameters of their chemical synthesis. Photorefractive polymers[47, 48, 49, 50] are a class of organic materials that possess charge generation, transport and trapping properties and exhibit an electrooptic effect. In general, different chemical species are responsible for these functions. Diffraction efficiencies as high as 90% have been demonstrated in photorefractive polymers. Most materials require the application of a strong electric field across the film during recording and readout. For example, the diffraction efficiency of photorefractive holograms recorded in PMMA:DTNBI:C_{60} increases as the fourth power of the externally applied electric field[51]. The electric field can play different roles, including enhancing hole mobility, increasing the efficiency of charge generation in chemically sensitized media, and aligning molecules responsible for the electrooptic effect[49].

Holographic photopolymers are another promising family of materials for write-once-read-many (times) (WORM) memories. A chemical reaction is catalyzed by exposure to light, resulting in a permanent index change. Photopolymers are available in formulations optimized for transmission or reflection holograms, and with sensitizer dyes that respond to different wavelengths. DuPont's HRF-150 has been extensively characterized at Caltech and elsewhere[9, 52, 53]. The material is available in 38 μm and 100 μm thicknesses on a Mylar substrate, and is suitable for high-density data recording[54]. A density of 100 bits/μm^2 has recently been demonstrated by Allen Pu at Caltech. One difficulty associated with photopolymers such as HRF-150 is a lateral shrinkage of the film during the curing process. Shrinkage impairs Bragg matching during readout. However, recording geometries have been devised to minimize the

impact of this effect.

A common problem affects both photorefractive polymers and holographic photopolymers such as HRF-150. It is usually difficult to prepare thick (around 1 cm) samples with sufficient optical quality. The limited available thickness reduces the data capacity of the samples. However, a candidate solution to this problem exists and is being investigated. It is possible to cascade several holographic films, separated by buffer layers, to form *stratified volume holographic optical elements*[55] (SVHOE). SVHOEs employing photopolymer[56] and photorefractive polymer[57, 51] films have been demonstrated.

Until recently, it has been difficult to compare the performance of recording materials because of a lack of standardization of testing procedures, conditions and figures of merit. The situation has improved, thanks to the introduction[41] and the widespread adoption of $M/\#$ as a figure of merit. Knowledge of $M/\#$ readily provides an estimate of the achievable diffraction efficiency (see equation 1.8). $M/\#$ can be easily estimated by measuring the initial slope of the recording curve and the hologram decay time constant (the slope of the logarithm of the decay curve).

A common reference ground for the evaluation of holographic materials was provided by the Photorefractive Information Storage Materials (PRISM) effort[58, 59], funded by the Defense Advanced Research Projects Agency (DARPA). The PRISM consortium, which includes IBM's Holographic Optical Storage Team (HOST), built a precision tester that can consistently assess the properties of recording materials at different wavelengths and using different recording geometries and multiplexing methods.

1.2.4 Laser diodes

The performance of holographic systems depends on the availability of powerful lasers operating at a wavelength at which good recording materials are available. It is generally recognized that the success of holographic memories in mass markets would be helped by powerful, inexpensive electrically pumped laser diodes. Compared to gas

lasers and optically pumped lasers, laser diodes are compact and inexpensive, and their cooling requirements are in general simpler. Fortunately, optical information storage and processing systems, and holographic information storage and processing systems in particular, can leverage the development of semiconductor lasers driven by the two applications of optics in information technologies described in Section 1.1.1: optical storage in compact discs and optical communications using optical fiber. Most commercially available laser diodes are based on III-V semiconductors and emit radiation in the red and the near infrared. This operating wavelength range is almost ideal for fiber communication systems because it corresponds to a region of low losses and low dispersion in silica fiber. However, shorter wavelengths would make laser diodes compatible with more inexpensive and well-characterized recording materials, such as $LiNbO_3$. Fortunately, compact discs are also pushing in the direction of shorter wavelengths, because the areal storage density of disks increases as the wavelength decreases. For example, the DVD standard employs 635 nm–650 nm laser diodes, down from 780 nm–790 nm in CD readers. A number of research institutes and corporations are developing blue, violet and even ultraviolet lasers, based on III-nitride semiconductors (e.g., GaN, InN, AlN, InGaN) and other materials[60, 61, 62]. Room-temperature operation of violet and ultraviolet laser diodes has been reported[60].

While edge-emitting diodes are common, vertical cavity surface-emitting lasers (VCSEL) emit light perpendicularly to the wafer, and are suitable for the fabrication of large laser arrays that are easily coupled to fiber bundles[63]. Moreover, they can be integrated monolithically with gallium arsenide circuits.

1.2.5 Other components

Several other needed component technologies have also benefited from significant recent advances. The fierce competition in the compact disc reader market has spawned a dramatic price reduction for injection-molded components (lenses and beam splitters, in particular) and mass-produced optical assemblies. Beam splitters and lenses cost only a few dollars apiece in volume.

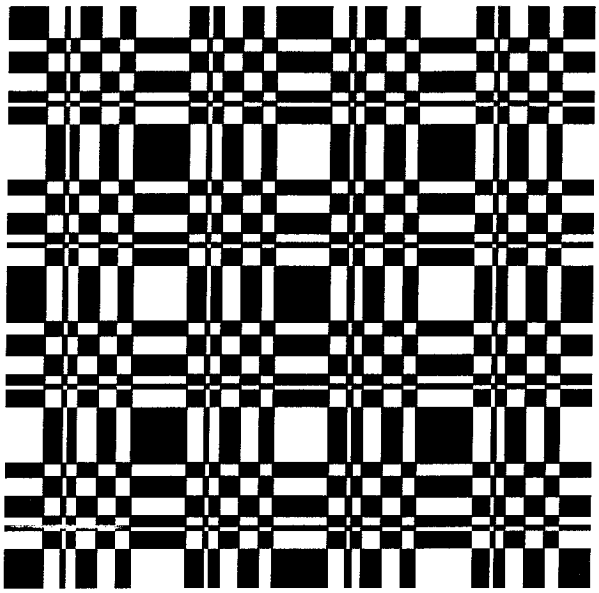


Figure 1.6: A Dammann grating designed to produce a 10 $\times$ 10 spot array.

Lithography techniques developed for the semiconductor industry have fostered the availability of high-quality diffractive optical elements (DOE), as did the development of efficient optimization algorithms and the accessibility of powerful computers. Such elements are used as lightweight lenses, planar holograms, array illuminators, beam shaping devices, etc. For example, Figure 1.6 shows a two-dimensional even-orders-missing Dammann grating[64] (a type of array illuminator) designed by the author to produce a 10 $\times$ 10 array of spots in its Fourier plane. Such gratings are used, for example, to illuminate optically active areas of smart pixel arrays. They can lead to substantial light-efficiency improvements when illuminating smart pixel arrays with low fill factor.

Although most DOEs have a fixed function, some are electronically controlled or programmable, and use, e.g., liquid crystals to modify their properties[65, 66]. One important example of such programmable DOEs are liquid crystal beam steerers[67, 68, 69]. We will see in Section 1.3 that they constitute compact, lightweight addressing devices in angularly multiplexed holographic systems that are free of most of the shortcomings associated with alternative devices.

1.3 System architectures

Laboratory prototypes based on the general architectures shown in Figures 1.1 and 1.2 have successfully proved the potential of holographic information storage and processing systems and subsystems for high-capacity data storage[70, 71, 72, 22], fast access[73] and complex pattern recognition tasks such as face recognition[74]. A logical next step is to try to simplify these architectures, integrate them into compact and more robust modules, and reduce their cost, in order to improve their manufacturability and ultimately their competitiveness. For example, consider the memory architecture shown in Figure 1.1. We will show that it is possible to

- remove half of the signal beam path;
- merge the SLM and detector array into a single device;
- relax the alignment requirements;
- eliminate the spacing between the components and assemble them in a compact module;
- provide a simple and elegant means to sustain dynamic holograms (see §1.2.3).

These improvements are based on a holographic module architecture that we started developing in 1993. The architecture is based on the following elements:

- conjugate readout;
- an OEIC combining SLM, detector array and memory functions, called *Dynamic Hologram Refresher*, or DHR for short.

Figure 1.7 shows such a compact architecture, employing angular multiplexing. In Chapter 5, we argue that the applicability of our integrative approach is general and we outline module architectures based on several other multiplexing methods. The approach also leads to a compact, elegant implementation of a combination correlator/associative memory, also described in Chapter 5.

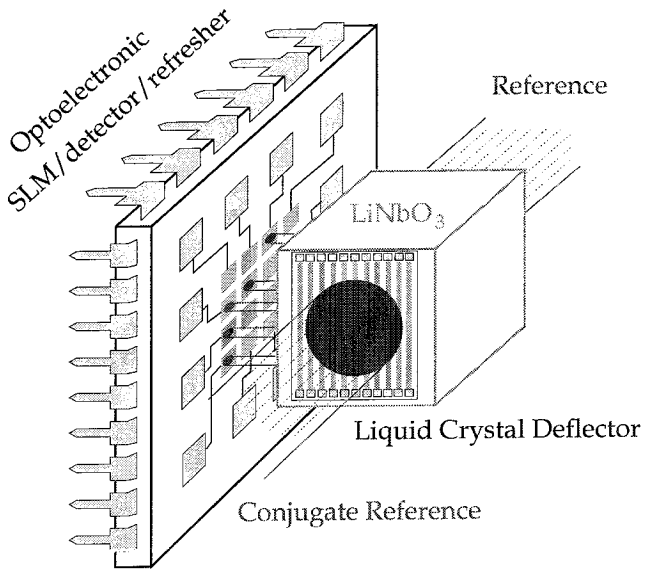


Figure 1.7: Compact angularly multiplexed holographic memory module.

The module is comprised of a photorefractive crystal, such as lithium niobate, in which holograms are stored, a pair of liquid crystal beam steerers (one of which is hidden behind the crystal) that are responsible for angularly multiplexing holograms in the crystal, and an optoelectronic integrated circuit (OEIC), the *Dynamic Hologram Refresher* or DHR, that merges the functions of a reflective SLM (for recording holograms) and a detector array (for readout). The DHR is also instrumental in sustaining holograms by periodic refreshing.

Holograms are written by letting the signal beam modulated by the DHR, **S**, interfere with the reference beam coming from the back of the crystal, **R**. The interactions between the beams and the gratings are represented in the **k**-sphere (momentum-space) diagrams of Figure 1.8. In general, the signal beam has a non-zero bandwidth because of the information imprinted on it by the DHR. This is represented by the dashed lines delimiting a signal "cone" in Figure 1.8. Therefore, the wave-vectors of the holographic grating fan out from the tip of the reference wave-vector on the circle, to reach all the spectral[16] components of the signal beam (see Figure 1.8a).

[16]We refer here to the Fourier spectrum of the information imprinted on the signal beam, rather than a spectrum of wavelengths; the beams are assumed to be nearly monochromatic (i.e., to have a

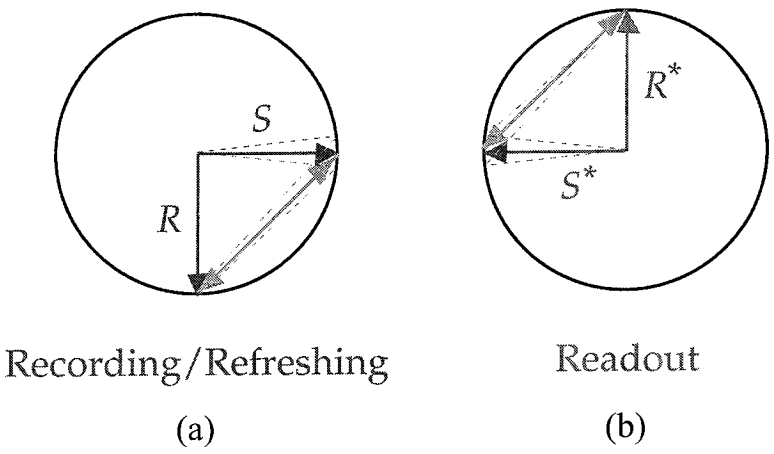


Figure 1.8: k-sphere representation of recording (a) and conjugate readout (b) interactions.

Holograms are read out by illuminating the crystal from the front with a counter-propagating reference beam, which is the conjugate $\mathbf{R}^*$ of the reference wave $\mathbf{R}$ employed to write holograms (see Figure 1.8b). $\mathbf{R}^*$ is Bragg-matched to all the components of the holographic grating[17]. The resulting conjugate reconstruction self-focuses back on the DHR, where it is sensed and processed.

The concept of conjugate readout in holographic systems, which is related to self-pumped phase conjugation[75], was first publicly disclosed at the OSA annual meeting in 1995[76]. Accuwave Corp. (Santa Monica, CA) recently demonstrated it in wavelength-multiplexed memory modules[77]. Chapter 6 will describe experimental demonstrations of integrated holographic memory modules based on conjugate readout.

The important advantages of the integrative approach are best appraised when the compact module of Figure 1.7 is compared to the conventional angularly multiplexed memory shown in Figure 1.1. In the latter system, holograms are stored in the recording medium by the interference of a signal beam coming from the upper left corner

narrow optical bandwidth). A Fourier component of spatial frequency u on the DHR gives rise to a plane wave possessing a transverse momentum component equal to $2\pi u$; the longitudinal momentum component is reduced accordingly, to maintain the overall momentum of the beam[2].

[17]In other words, the tip of the vectorial sum of $\mathbf{R}^*$ (with its origin in the center of the sphere of radius $2\pi/\lambda$) and any of the grating wave-vectors lies on the sphere.

with a reference beam from the upper right corner. A transmissive SLM imprints data on the signal beam, while a motorized rotating mirror addresses individual data pages within the memory. During readout, the reconstructed signal beam propagates away from the SLM, toward a detector array (for example, a CCD camera) located in the lower right corner of the figure. We note the following points.

- Lenses are required to image the reconstructed data page on the detector array. The required distance between the lenses, the SLM and the detector array imposes a minimum size on the system. Lens aberrations are also a major problem when large pages of data are stored.

- The detector array must be aligned to within a fraction of the pixel size with the SLM.

- The motorized mirror is big and heavy, and suffers from a high power dissipation, low speed and backlash (repeatability) problems.

The compact system architecture addresses all these issues and provides additional benefits as well. The module shown in Figure 1.7 contains no lenses. Conjugate readout eliminates the need for imaging optics, since the reconstruction retraces the path and phase profile of the signal beam during recording. This allows all the components to be assembled in a single compact structure. If a single OEIC containing modulators and detectors at each pixel location (such as the DHR chip) is employed, alignment is guaranteed by design. Furthermore, the page addressing devices we propose to employ (e.g., the liquid crystal beam steerer in Figure 1.7) are free of the problems associated with conventional mechanical addressing systems (vibrations, backlash, low speed, high power consumption) and with acousto-optic deflectors (wavelength shift, high-voltage operation). Finally, because phase conjugation undoes aberrations in the signal path, phase-conjugate reconstructions possess high fidelity.

1.4 Outline of the thesis

The need for optical output devices added to silicon OEICs with a minimum of post-processing was explained in Sections 1.2.1 and 1.2.2. Chapter 2 will describe hybrid-aligned nematicliquid crystal modulators fabricated on VLSI die (HAN-on-VLSI modulators). The modulators provide true analog amplitude or phase modulation at low voltage and at video rates.

Chapter 3 will describe a dynamic hologram refresher OEIC designed, tested and employed in laboratory prototypes of modular holographic memory at Caltech. The semiconductor die was fabricated in Orbit Semiconductor's 2.0 μm double-metal n-well process through MOSIS[78]. HAN-on-VLSI modulators fabricated in-house provide optical outputs. The chapter then describes a very compact next-generation DHR pixel design, with a thorough noise analysis.

In Chapter 4, we present two optoelectronic neural arrays with sigmoid and "bump" responses. One array employs Lucent Technologies' Hybrid-SEED process; the other uses light-emitting diodes grown by molecular beam epitaxy on gallium arsenide die fabricated in Vitesse Semiconductor's H-GaAs-3 process available through MOSIS.

Chapter 5 describes practical implementations of the integrated holographic memory module shown in Figure 1.7. An analysis leading to cost and density projections for integrated modular holographic memory is presented. Integrated module architectures based on wavelength, phase-code and shift multiplexing are also described.

Finally, the results obtained from laboratory prototypes of integrated modular holographic memory are presented in Chapter 6. The prototypes demonstrated the following functions:

- lens-less conjugate readout using a counter-propagating reference beam;
- read/write functions using a single OEIC as SLM and detector array;
- sustainment of a dynamic hologram over 50 refresh/decay cycles.

Chapter 2 Liquid-crystal modulators

The topic of this Chapter is liquid-crystal modulators fabricated on VLSI die. Such reflective modulators provide low-power amplitude or phase modulation of readout beams at voltages compatible with advanced CMOS processes. Most types of liquid-crystal-on-silicon (LCOS) modulators can be assembled on conventionally fabricated VLSI die with a minimum of post-processing. This Chapter emphasizes hybrid-aligned nematic liquid-crystal modulators fabricated on VLSI die[79] (HAN-on-VLSI).

We begin with a brief introduction to LCOS modulators, and HAN-on-VLSI modulators in particular, in §2.1. In §2.2, we present some preliminary results obtained with hybrid-aligned nematic (HAN) cells using glass substrates. The fabrication procedure of HAN-on-VLSI modulators is described in §2.3; experimental results are presented. The procedure relies on a chemical surface-coupling agent to align the liquid crystal on the surface of the VLSI die. An alternative approach, based upon the microstructure of the surface of the die, is discussed in §2.4.

2.1 Liquid-crystal-on-silicon modulators

The development of highly integrated spatial light modulators and smart pixel arrays based on VLSI circuitry and liquid crystal devices has made significant progress over the past few years[80, 81, 82]. Figure 2.1 shows a cross-section view of a generic LCOS OEIC. A thin layer of liquid crystal is sandwiched between the VLSI die and a transparent cover plate. The cover plate is maintained at a small distance (a few μm) above the die by microscopic spacers. The bottom face of the cover plate is coated with a transparent electrode used to electrically drive the liquid crystal, and usually also with a thin layer of a surface-coupling agent that aligns the liquid crystal molecules. Figure 2.1 illustrates that LCOS OEICs combine optical detection, information processing, and optical modulation functions, respectively represented in the Figure by a

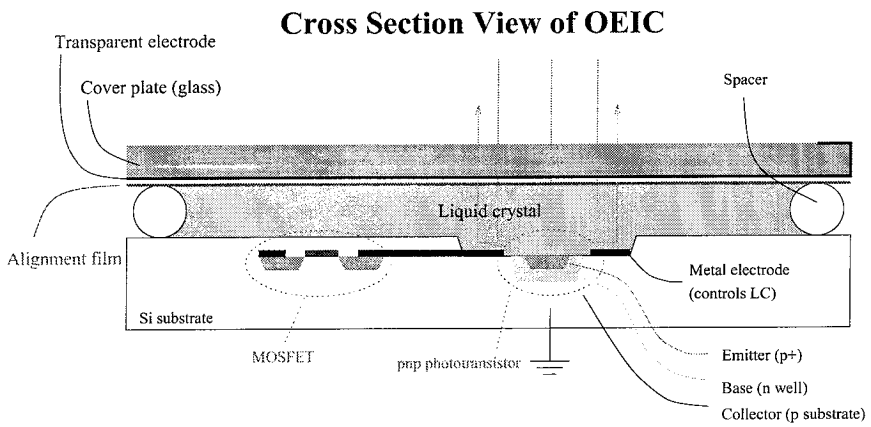


Figure 2.1: Cross-section view of a generic LCOS OEIC.

phototransistor, a MOSFET and the liquid-crystal modulator structure.

While some early work was conducted with nematic liquid crystals[83], most current efforts involve ferroelectric chiral smectic C liquid crystals in the surface-stabilized geometry[84, 85, 86, 87, 88] that provide fast switching at low voltages. We demonstrate an alternative modulator structure that employs nematic liquid crystals in the *hybrid-aligned nematic* (HAN) configuration, which provides analog amplitude or phase modulation of an incident monochromatic readout beam. Prototype modulators fabricated on commercially processed VLSI devices have contrast ratios in excess of 18:1, and can operate at video rates.

Surface-stabilized ferroelectric liquid crystal modulators require a very small cell gap, on the order of 1 μm to 2 μm, in order to ensure bistability and to obtain a zero-order half-wave retardation in one of the two stable states. This small cell gap makes the device very sensitive to the surface topology of the VLSI die, and a planarization treatment is generally required to improve contrast and uniformity[81]. In the case of nematic liquid crystals with larger cell gaps, the topmost layer of the integrated circuit, which acts as a bottom substrate for the liquid crystal cell, must provide strong anchoring of the neighboring nematic. While various surface treatments can provide strong homogeneous alignment, the deposition of a silane compound inducing homeotropic alignment provides strong anchoring at the bottom substrate, while requiring a

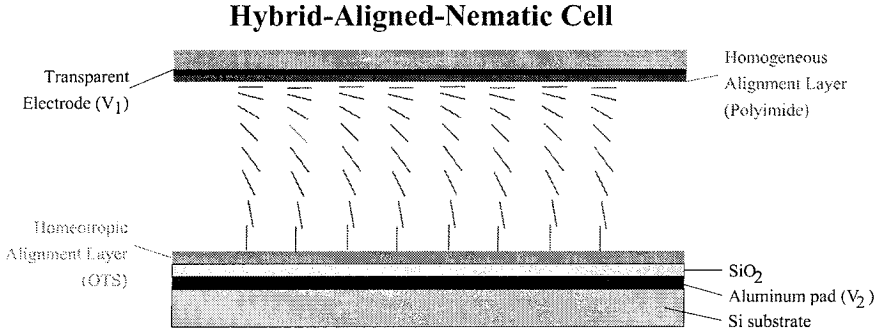


Figure 2.2: Schematic illustration of a reflective hybrid-aligned nematic cell fabricated on the surface of a VLSI die or wafer.

minimum of post-processing to be performed on the fabricated semiconductor die or wafer.

2.2 Hybrid-aligned nematic modulators

In hybrid-aligned nematic liquid crystal modulators, one substrate (the *cover plate*) induces strong homogeneous alignment, whereas the other causes strong homeotropic alignment. By virtue of bulk elasticity, the orientation of the nematic director[1] undergoes a smooth rotation from one substrate to the other[89], causing a splay-bend conformation in the nematic film, as illustrated in Figure 2.2. When no electric field is applied across the cell, the locally averaged orientation of the molecules varies smoothly from a vertical state on the surface of the die to a horizontal state close to the cover plate. Using the Frank elasticity theory, assuming strong anchoring at both substrates and using the one-constant approximation ($K_1 \approx K_3$), it can be shown that the angle (in radians) between the nematic director and the vertical axis (perpendicular to the substrates), $\phi(z)$, is approximately a linear function of depth within the cell:

$$\frac{1}{d}\left(z+\frac{d}{2}\right) = \frac{\int_0^{\phi(z)} \sqrt{K(x)}dx}{\int_0^{\frac{\pi}{2}} \sqrt{K(x)}dx}, \tag{2.1}$$

[1]The *director* is defined as a unit vector pointing in the locally averaged direction of the long axis of the oblong liquid crystal molecules. It usually coincides with the optical axis.

where d is the cell gap (the distance between the two substrates), $z = -d/2$ on the surface of the chip, and

$$K(\phi) \equiv K_3 \left(1 - h \sin^2(\phi)\right), \tag{2.2}$$

where

$$h \equiv 1 - \frac{K_1}{K_3}, \tag{2.3}$$

and K_1 and K_3 are the splay and bend elastic coefficients of the mesophase, respectively. In the simplifying so-called *one-constant approximation*, where $K_1 \approx K_3$, the angle between the director and the vertical is approximately a linear function of distance:

$$\phi(z) \approx \frac{\pi}{2d}\left(z + \frac{d}{2}\right). \tag{2.4}$$

Under these conditions, assuming illumination by a normally incident beam, the total one-way phase retardation between the two eigenmodes of the cell is

$$\Delta\phi \approx \frac{2\pi}{\lambda} n_o d \left(\frac{2}{\pi}\int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - R\sin^2\phi}} - 1\right), \tag{2.5}$$

where $R \equiv 1 - n_o^2/n_e^2$, n_o and n_e are respectively the ordinary and extraordinary indices of refraction of the uniaxial nematic liquid crystal, and λ is the wavelength of the incident optical beam in vacuum.

When an electric field is applied, the director tends to align with it. In the high-field limit, the director is perpendicular to the substrates throughout the cell, except for a thin film in the immediate vicinity of the cover plate, and there is practically no phase retardation between the eigenmodes of a normally incident plane wave. The effective index of refraction seen by light polarized along the buffing direction of the cover plate can be continuously changed by applying analog voltages across the cell. Given a suitable input light polarization, this phase modulation can be converted into amplitude modulation with the help of an analyzer.

The electrooptic response of a typical HAN sample is shown Figure 2.3. The device, which had glass substrates and a cell gap of 10 μm, was filled with the nematic

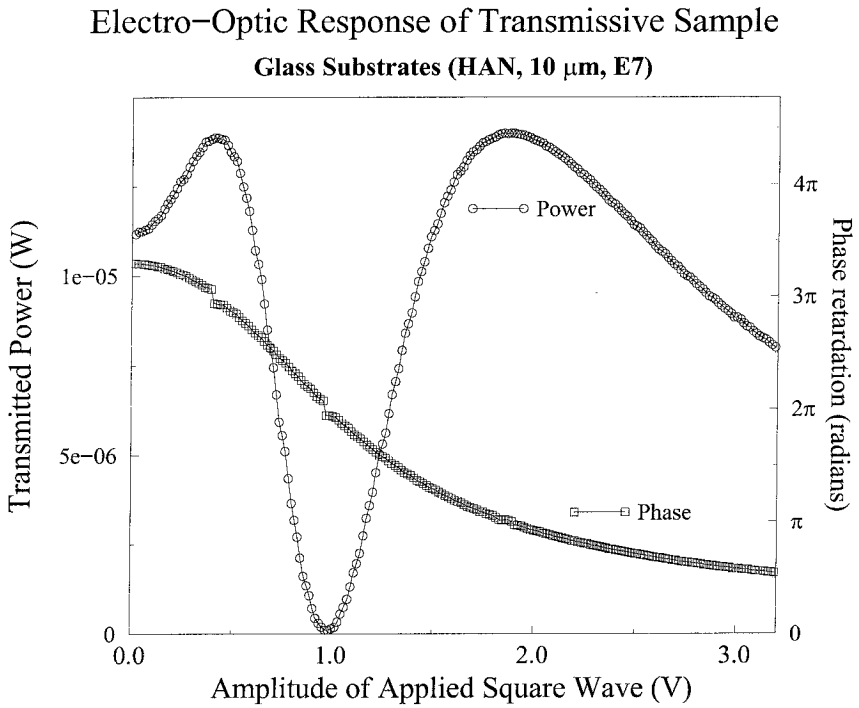


Figure 2.3: Electrooptic characteristics of a transmissive HAN cell.

E7 from Merck, placed between crossed polarizers, illuminated by a collimated laser beam at 632.8 nm, and driven by a 1 kHz square wave of variable amplitude. The surface coupling agents inducing homogeneous and homeotropic alignment were rubbed polyimide and octadecyltriethoxysilane (OTS), respectively. As expected, the transmission asymptotically goes to zero in the high-voltage regime, where the director (and hence the optic axis) is nearly normal to the substrates throughout the cell and the phase retardation between the eigenmodes excited by the normally incident beam is nearly zero. The phase retardation increases as the amplitude of the applied square wave decreases, until it reaches a maximum value determined by the cell gap, the properties of the nematic (indices of refraction, elastic constants) and the anchoring strength at the substrates. The phase retardation of the undriven device is seen to be 3.3 radians, in excellent agreement with the value (3.4 radians) predicted by equation 2.5, with the appropriate indices of refraction[90] $n_o = 1.5211$ and $n_e = 1.7464$, and assuming strong anchoring at both substrates. Note that unlike parallel-rubbed cells, HAN devices have no threshold voltage for the onset of an electrooptic response.

2.3 HAN-on-VLSI modulators

We have also fabricated reflective HAN modulators on semiconductor die processed by commercial foundries. In particular, our fabrication method is compatible with the Orbit Semiconductor 2.0 μm silicon-gate complementary metal-oxide-semiconductor (CMOS) process available through MOSIS[78].

Figure 2.4 schematically shows the detailed structure of a HAN-on-VLSI device. Metal electrodes made of one of the metallization levels provided by the semiconductor process are used to apply voltages locally across the liquid crystal cell, and as optical mirrors. They are overlaid by a silicon dioxide protective overcoat, which is also a native layer of the VLSI process, and which helps octadecyltriethoxysilane (OTS), the surface coupling agent inducing homeotropic alignment in the nematic film, to bind to the surface of the die. An OTS solution (2% mass in ethanol) is spun on the VLSI die or wafer at 3000 RPM for 40 s, and then baked at 200°C for 30 minutes. This

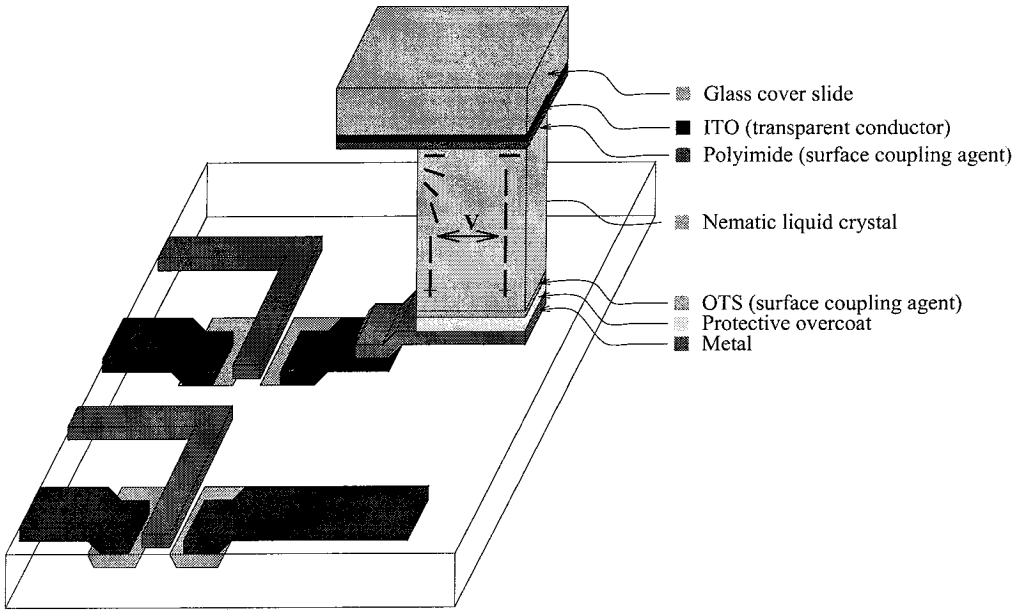


Figure 2.4: Structure of a HAN-on-VLSI modulator.

low-temperature operation does not affect the performance of the electronic circuitry. The glass cover plate is coated with a 20 $\Omega/\square$ indium-tin-oxide (ITO) transparent electrode and with uniaxially rubbed polyimide. Aluminum is evaporated on one edge of the cover plate to allow a wire to make a good contact to the ITO backplane. The cover plate is affixed to the surface of the chip, at a microscopic distance set by small drops of a mixture of chopped glass fibers and Norland 61 UV-cured optical adhesive deposited on the periphery of the active area of the device. The resulting cavity is filled with Merck E7 in the isotropic phase. The mesogenic substance enters the nematic phase as the device slowly cools down. Figure 2.5 shows a complete OEIC with a HAN-on-VLSI modulator assembled in our laboratory.

Figure 2.6 shows the electrooptic response of a device employing as a lower substrate an integrated circuit fabricated in the Orbit 2.0 μm n-well CMOS process available through MOSIS. The cell gap is 5 μm, and the cavity is filled with Merck E7. Crossed polarizers were used. The response of an area containing approximately six externally driven pixels was measured. Figure 2.7 shows rows of pixels at maximum and minimum reflection settings. Voltage-dependent reflectance is also observed in the

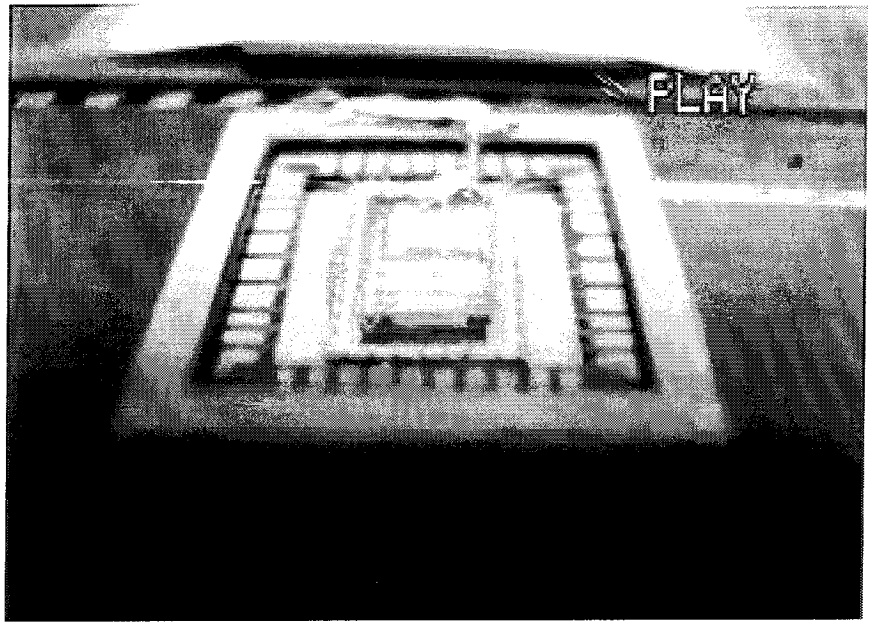


Figure 2.5: Optoelectronic integrated circuit with a complete HAN-on-VLSI modulator.

areas surrounding the pixels, due to the presence of conductors at various potentials and depths within the silicon dioxide. The total phase retardation of the undriven device, 11.4 radians, is within 7.6% of the value predicted by equation 2.5, where d was set to 10 μm to account for the two passes of the beam through the modulator. Note that the entire amplitude modulation range is spanned by applying voltages less than 2.2 V across the modulator electrodes. One of the reasons why the required voltage is higher than in the case of the device with glass substrates is the voltage drop across the protective overcoat (Figure 2.4). Imaged contrast ratios in excess of 18 : 1 have been obtained in HAN-on-VLSI prototypes, without any planarization treatment on the die. The first level of metallization was employed to form the pixel electrodes, since the second-level metal of the Orbit 2.0 μm processes is known to have a poorer optical quality.

Figure 2.8 shows the electrooptic responses of widely separated areas of the same device (sample 126) and of another device (sample 127) prepared using the same procedure. Note that the voltages corresponding to the first reflectance maximum and

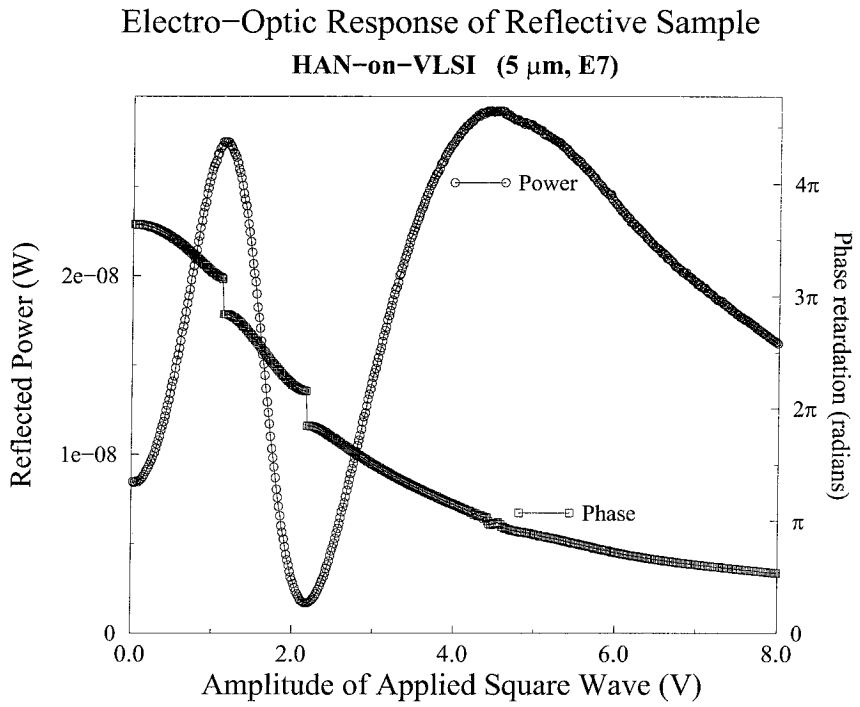


Figure 2.6: Electrooptic response of a HAN-on-VLSI device.

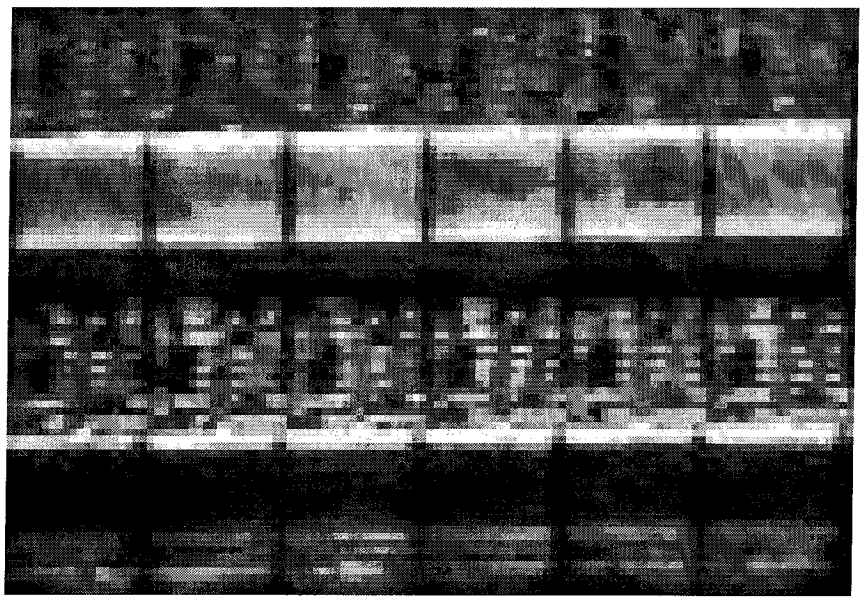


Figure 2.7: Rows of six pixels at maximum (top) and minimum (bottom) reflection settings.

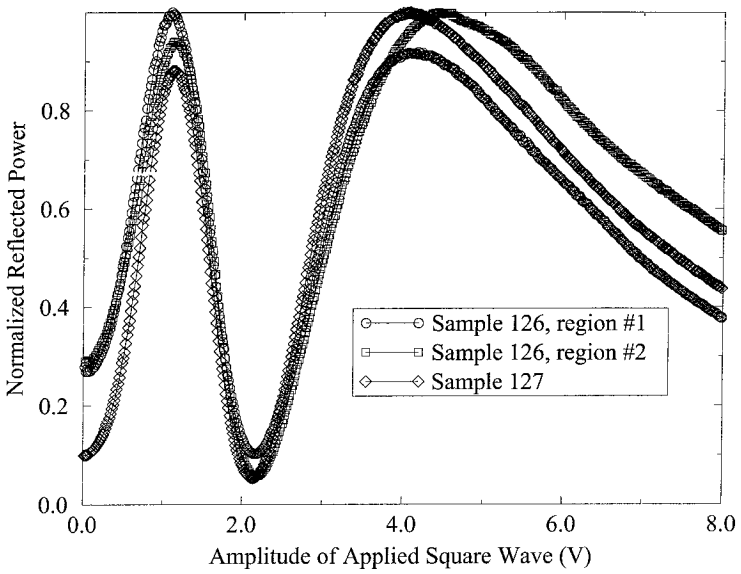


Figure 2.8: Electrooptic responses of different areas of the same device (sample 126) and of another device processed using the same procedure (sample 127).

minimum of the three curves are very close (within approximately 30 mV). This is an indication of the reliability and uniformity of the process.

The rise and fall times of the HAN-on-VLSI device whose electrooptic response is shown in Figure 2.6, with $V_{on} = 4.42$ V and $V_{off} = 2.17$ V, were 12 ms and 24.6 ms, respectively. Note that this allows the device to perform video-rate (30 Hz) modulation. If lower driving voltages are desired, the reflectance maximum at 1.12 V can be employed instead of that at 4.42 V; the resulting switching time then increases to 40 ms. Faster operation can be obtained with lower-viscosity liquid crystals and with reduced cell gaps.

Additional data characterizing HAN-on-VLSI modulators fabricated on DHR OE-ICs is presented in Chapters 3 and 6. §3.2.4, in particular, characterizes the reflectance distributions of on and off pixels and gives contrast-ratio statistics.

2.4 Alignment of liquid crystals by micropores

We saw in §2.3 that homeotropic alignment can be obtained by coating the die with a chemical surface coupling agent. Another approach is to alter the microstructure of the surface of the die. Most VLSI processes provide means of exposing the topmost metal level, which is often aluminum. The anodic oxidation of aluminum in a strongly acidic electrolyte results in the formation of a film of highly porous aluminum oxide, where the elongated microscopic pores have their long axis perpendicular to the substrate. A nematic placed in contact with such a porous medium can penetrate the pores, and be homeotropically oriented through a combination of bulk elasticity and surface interaction with the walls of the pores. The aligning effect of porous anodic aluminum oxide was experimentally demonstrated[91].

An alignment film is formed on a substrate coated with aluminum by anodizing the aluminum in an acidic environment. This forms a layer of porous aluminum oxide on the surface of the substrate[92, 93, 94, 95, 96]. The elongated pores have their long axis approximately normal to the surface of the substrate. The diameter of the pores is on the order of 20 nm; their depth can be up to several micrometers. The behavior of liquid crystals put in contact with such a substrate is consistent with homeotropic alignment. This is conceivably a result of a combination of the interaction between the nematic and the pore walls, and of the elastic behavior of nematic liquid crystals. Within the pores, the elastic deformation energy of the embedded liquid crystal is minimized when the director is parallel to the walls of the pores, or perpendicular to the surface of the substrate. This homeotropic alignment is carried to some extent to the bulk of the nematic film by bulk elasticity, as schematically illustrated in Figure 2.9.

The bulk of the nematic film is affected by both substrates. The other substrate can be of many types, such as, for example, a glass cover plate coated with a transparent electrode and a material that induces homogeneous alignment. Such an arrangement was used to demonstrate the aligning effect of porous anodized aluminum and the competition between the homogeneous and homeotropic alignments induced by the substrates. Several different samples were assembled with porous aluminum oxide

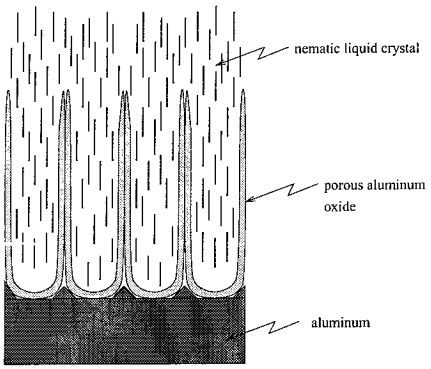


Figure 2.9: Schematic illustration of homeotropic alignment induced by porous aluminum oxide resulting from anodization of aluminum.

produced with different anodization times and voltages. One substrate was a glass slide with a 2 μm thick coating of evaporated aluminum, partially anodized in a sulfuric acid bath. The other substrate was coated with uniaxially rubbed polyimide. The electrooptic responses of the samples are shown in Figure 2.10. The substrate with porous anodized aluminum clearly induces homeotropic alignment, as evidenced by the phase retardation of the cells and by the absence of a threshold voltage[97]. If it did not, the configuration at rest would be undistorted (no splay nor bend), and the total phase retardation would be

$$\frac{4\pi d\Delta n}{\lambda} = 17.2 \text{ rad.}$$

In contrast, equation 2.5 predicts a phase retardation of 8.4 rad for a 10 μm reflective HAN cell with no applied voltage, in good agreement with the experimental data. The electrooptic responses appear to be relatively insensitive to the anodization parameters.

This alignment method, applied to the fabrication of OEICs, uses electrodes made of a metalization level native to the conventional VLSI fabrication process both optically, as mirrors, and electrically, to apply local fields across the nematic film, thereby defining picture elements (pixels). The aluminum of the topmost metalization level can usually be selectively exposed. It is thus possible to cover the entire surface of the

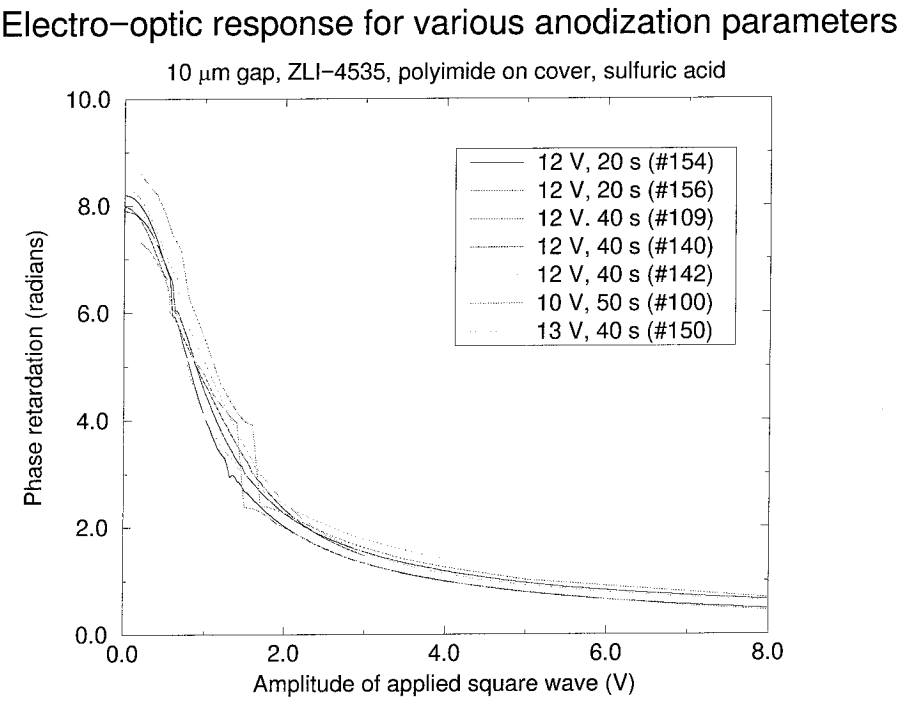


Figure 2.10: Electrooptic response of liquid-crystal cells with porous anodized aluminum.

die with a passivation layer, except for electrodes assigned to driving liquid crystals and bonding pads. The electrodes assigned to driving pixels can be selectively anodized by immersing the die (or an entire wafer, for that matter) in an acidic electrolyte and circulating an ionic current between it and a metallic electrode, thereby forming a layer of porous aluminum oxide on top of each pixel electrode. The aluminum pixel electrodes are only partially anodized, and the remaining aluminum can be employed for reflecting optical signals and for applying electric fields across the nematic film. The simple procedure can be performed on the die (or wafer) immediately after the conventional VLSI fabrication, and before the cover plate is affixed to it.

Chapter 3 The Dynamic Hologram Refresher

3.1 Introduction

One of the key elements of integrated modular holographic memory is the *Dynamic Hologram Refresher* optoelectronic integrated circuit (OEIC) that interfaces holographic modules to an electronic backplane[1], and whose pixels locally implement the following functions (see Figure 3.1):

1. optical detection;
2. memory;
3. optical modulation.

The optical detection function is used when reading out a data page from the holographic recording material. The modulator is used to imprint data on the signal beam during recording operations. The memory function is used to decouple the highly parallel (concurrent transfers of data pages whose individual capacity is on the order

[1] More details on this in §5.1.

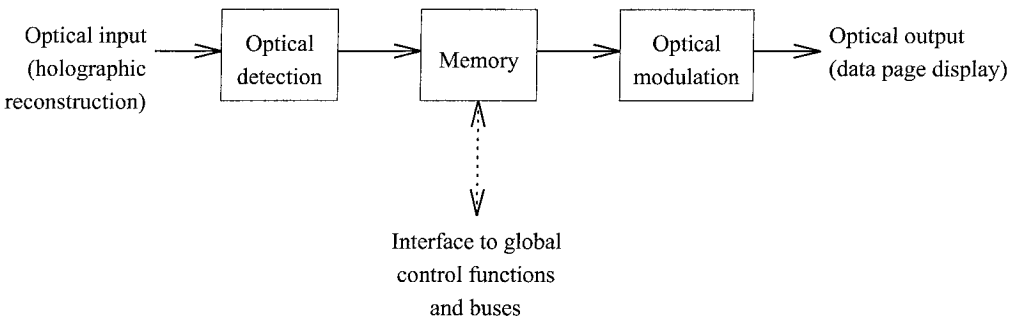


Figure 3.1: Functional diagram of a pixel.

of 2 Mb), but somewhat slow (fastest cycle time on the order of 1 ms), interaction with the recording material from the electronic transfers into and out of the holographic storage module, which possess little parallelism (8- to 128-bit-wide transfers, perhaps) but short cycle times (a few tens of nanoseconds). All three functions are also employed in sequence while refreshing dynamic holograms. A refreshment cycle proceeds as follows (see Figure 1.7). First, the desired data page is addressed and the conjugate (readout) reference illuminates the crystal from the front. The reconstructed data page is sensed and thresholded by the optical detectors within each pixel. The resulting binary data is transferred to the local memory bit, which in turn drives the pixel's modulator to the appropriate state. The signal beam is then shone on the OEIC, which modulates it with the sensed data page, as the conjugate reference is turned off and the main (recording) reference illuminates the crystal from the back. The interference of the signal beam with the main reference reinforces the decaying hologram.

Chapter 6 describes experiments that demonstrated conjugate readout, compact lens-less memory modules and sustained dynamic holograms. All the experimental data presented in Chapter 6 was obtained using the DHR-4 device, which is described in §3.2. §3.2 presents the architecture of the DHR-4 OEIC and its pixel circuitry, and an experimental characterization of its detectors and modulators.

The DHR-4 device uses static RAM and extensive signal buffering in the optical input train. As a result, DHR-4 pixels require only a minimal amount of relatively simple support circuitry in the periphery of the pixel array. However, the density of the array is low because of the high complexity (20 transistors) of the pixel circuit. This was the motivation behind the new *dynamic* DHR pixel designs. In §3.3, we show that all three functions required of a DHR pixel circuit (optical detection, optical modulation and memory) can be implemented with one photodiode, one capacitor and at most four MOSFETs. Moreover, all transistors can be native; therefore, the area overhead associated with wells is avoided. The operation of the new pixel designs is detailed, and their noise performance is estimated.

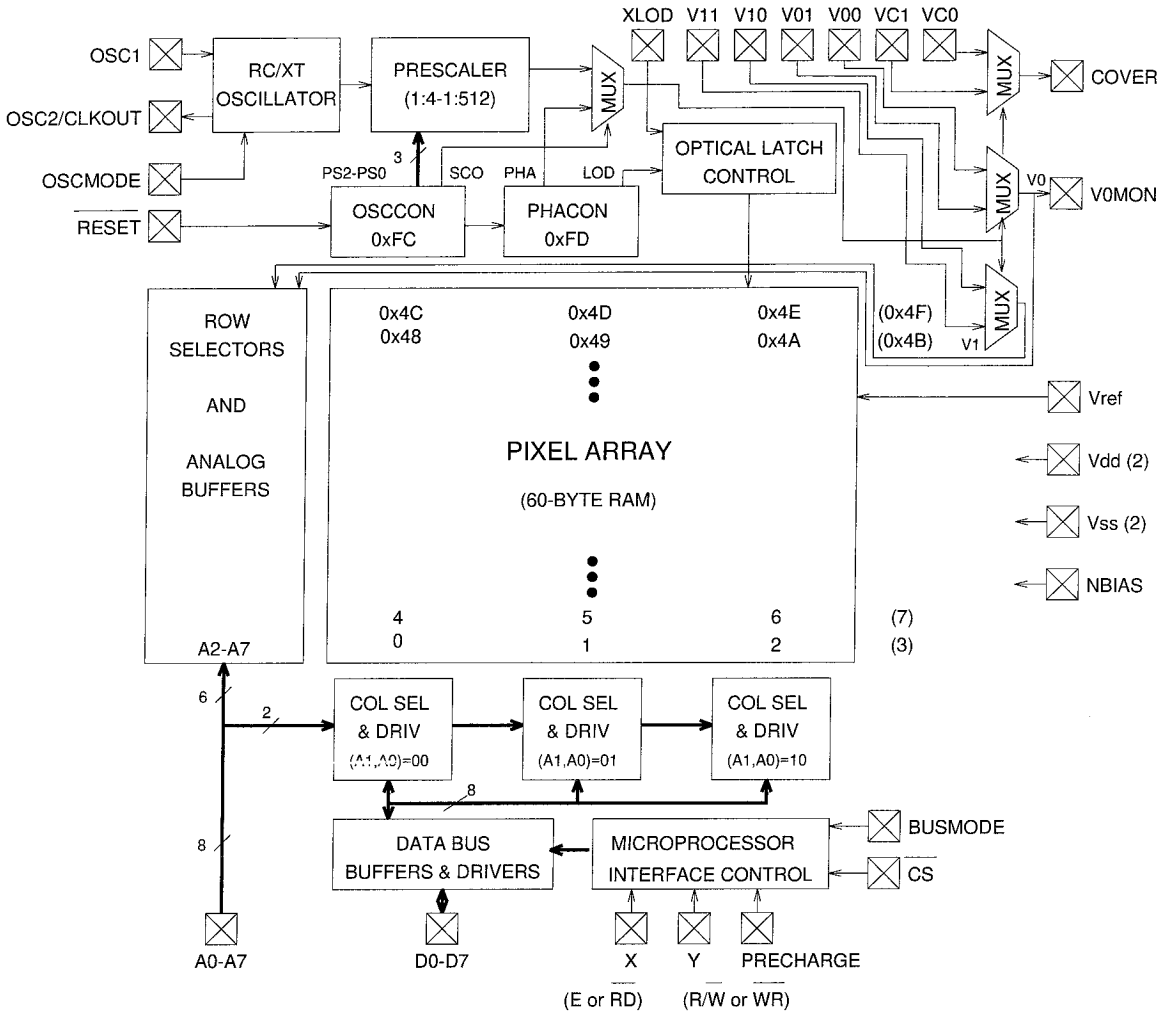


Figure 3.2: Functional diagram of the DHR integrated circuit.

3.2 DHR-4

3.2.1 Functional organization

The DHR-4 optoelectronic integrated circuit consists of an array of 480 pixels, arranged in a 20×24 matrix. It is intended to appear as a static RAM to a controlling microprocessor. The design emphasizes simple interfacing and control. Figure 3.2 shows a functional diagram of the DHR integrated circuit. The pixel array is accessed through an industry-standard 8-bit parallel microprocessor interface bus (D0–D7) that can be controlled using Intel- or Motorola-style control signals, as determined by the

BUSMODE input[2]. A group of eight pixels is associated with each valid address. Address bits A0 and A1 select eight columns of the pixel array at a time, while A2–A7 address one of the 20 rows, or the memory-mapped special-function registers OSCCON and PHACON. Each row contains 24 pixels, or three bytes. However, the two bits that select columns (A0 and A1) can address four different locations. This is why all addresses in which A0 = A1 =1 (locations number $3, 7, \ldots, 0x4B, 0x4F$)[3] are mapped to unimplemented memory bytes. Writing to these locations has no effect; reading from them returns invalid data.

Each pixel contains an optical detector, an optical modulator and a static memory bit (more details on the pixel circuit will be provided in §3.2.2). The modulator embedded within each pixel is on if and only if the corresponding memory circuit stores a 1. Data can be read from or written to the pixel's memory circuit through the data bus buffers and drivers. Optically presented data can also be memorized by pulsing the internal LOD *(Load Optical Data)* signal. This can be done either by applying a pulse to the XLOD pin, or by pulsing the LOD bit in special-function register PHACON (see Figure 3.4). The internal LOD signal is the logical OR of XLOD and PHACON<LOD>. Whenever it is active, each pixel's memory bit is driven by its corresponding optical detector.

The integrated circuit provides hardware to support the generation of DC-balanced liquid crystal waveforms, required to ensure a long modulator lifetime. Three analog multiplexers select the appropriate voltages for the cover plate (backplane) and "on" and "off" pixels (V_c, V_1 and V_0, respectively), depending on the active half-cycle of the clock signal Φ_{LC}. Table 3.1 shows the voltage assignments. DC balancing is achieved if the clock has a 50% duty cycle (this is always the case if the internal oscillator is used) and the following conditions hold:

$$V00 - VC0 = -(V01 - VC1) \tag{3.1}$$

[2]Intel-style bus control signals ($\overline{RD}$ and $\overline{WR}$) are used when BUSMODE = 0; Motorola-style signals (E and R/$\overline{W}$) control the buses when BUSMODE = 1. The active-low chip-select signal $\overline{CS}$ is employed in the same way in both modes.

[3]We use the same numerical notation as in the C programming language. The prefix "0x" denotes a hexadecimal number.

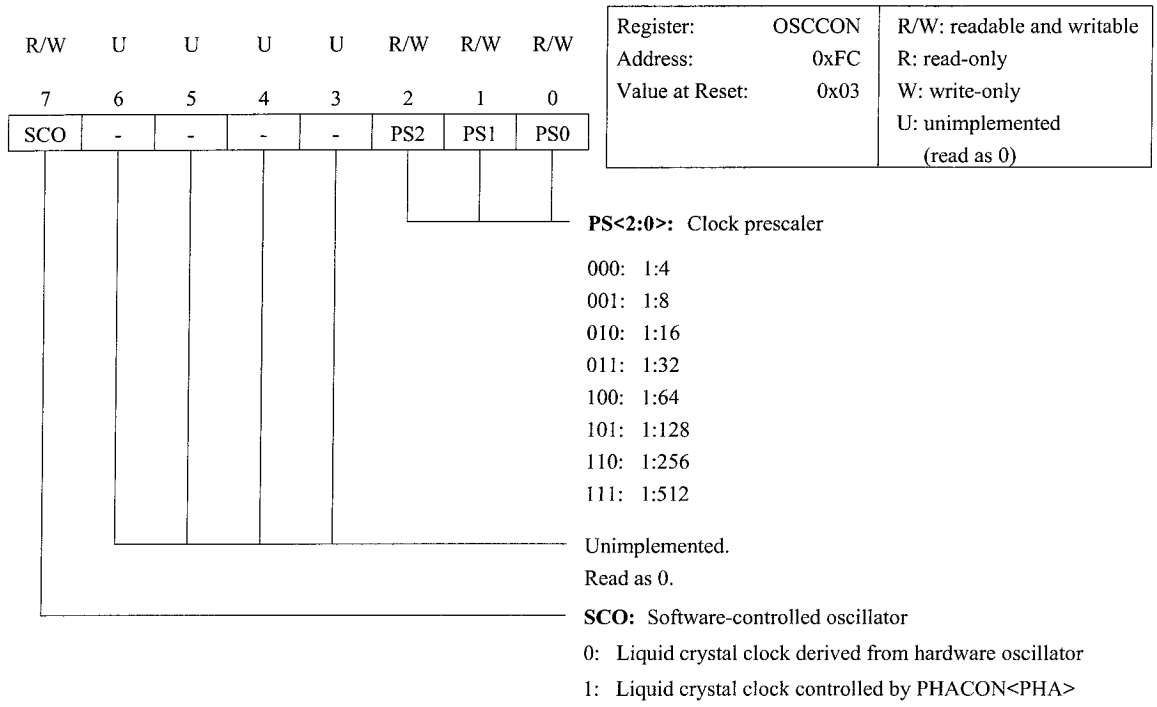


Figure 3.3: Bit assignment of special-function register OSCCON.

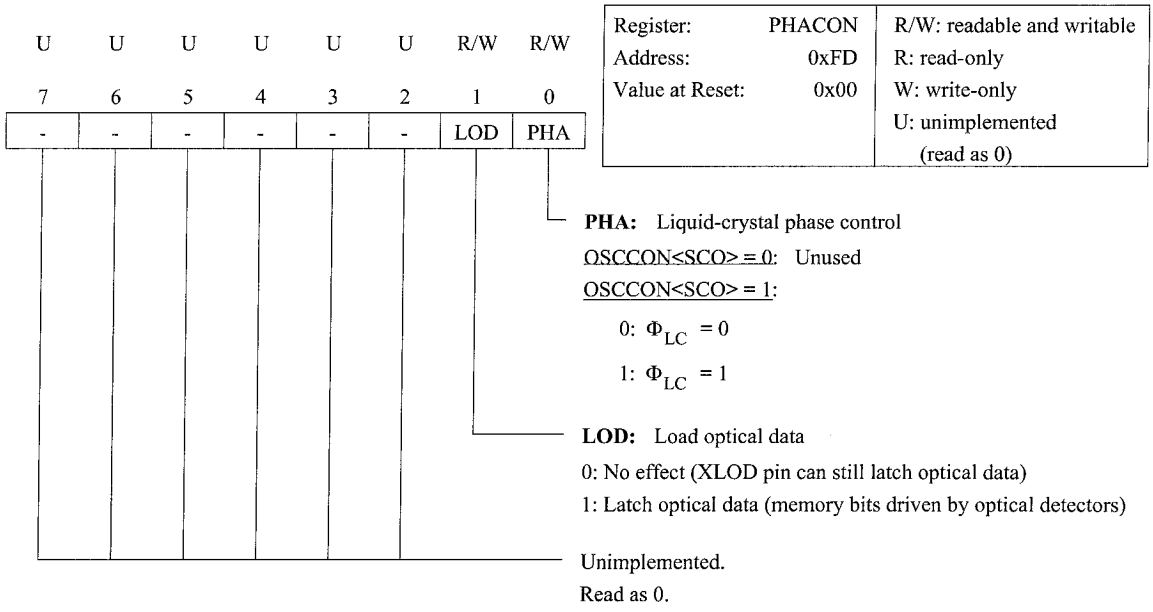


Figure 3.4: Bit assignment of special-function register PHACON.

Φ_{LC} (clock phase)	0	1
V_c (backplane)	VC0	VC1
V_0 (LC electrode of pixels storing a 0)	V00	V01
V_1 (LC electrode of pixels storing a 1)	V10	V11
$V_{LC,0}$ (voltage across LC in a pixel storing a 0)	V00 – VC0	V01 – VC1
$V_{LC,1}$ (voltage across LC in a pixel storing a 1)	V10 – VC0	V11 – VC1

Table 3.1: Operation of the multiplexers providing the signals to drive the liquid crystal modulators.

$$\text{V10} - \text{VC0} = -(\text{V11} - \text{VC1}). \tag{3.2}$$

One way to satisfy the above conditions is to assign the voltages in such a way that the signals used in the low phase of Φ_{LC} are the complements (with respect to the supply rail V_{dd}) of the signals employed in the high phase of Φ_{LC}:

$$\text{VC1} = V_{dd} - \text{VC0} \tag{3.3}$$

$$\text{V01} = V_{dd} - \text{V00} \tag{3.4}$$

$$\text{V11} = V_{dd} - \text{V10}. \tag{3.5}$$

If the liquid crystal voltages are generated by a unipolar digital-to-analog converter (DAC), equations 3.3–3.5 are satisfied if the binary codes assigned to a pair of liquid crystal voltages (e.g., VC0 and VC1) add up to the full-scale code. For example, with an 8-bit converter (the full-scale code is 0xFF), if the binary code of VC0 is 0x23, that of VC1 should be 0xDC.

The analog multiplexers consist of CMOS transmission gates and buffer amplifiers. The two signals distributed to the pixel array, V_0 and V_1, are buffered at each row (in the box on the left side of the pixel array in Figure 3.2), to avoid fluctuations caused by excessive loading of the multiplexers.

The liquid-crystal clock, Φ_{LC}, is obtained from one of two sources. If OSCCON<SCO> = 1 (see Figure 3.3), Φ_{LC} = PHACON<PHA>. Φ_{LC} is then under software control; its state is changed by writing to special-function register PHACON. If OSCCON<SCO> = 0,

Φ_{LC} is derived from the output of the on-chip hardware oscillator[4] [5]. The output of the oscillator is fed to a prescaler consisting of a string of flip-flops with multiple taps. Each flip-flop divides the input clock frequency by two. An 8:1 multiplexer selects a frequency division ratio under control of OSCCON<PS2:PS1> (see Figure 3.3). The ratio can be any power of two between 4 and 512, inclusively. An external clock can be used instead of a crystal or an RC network by connecting it to the OSC1 pin, by selecting the crystal mode (OSCMODE = 0), and by leaving the OSC2 output unconnected.

The active-low Schmitt-triggered $\overline{\text{RESET}}$ input initializes the internal circuits to a known state. It can be directly connected to an external RC network to automatically RESET the DHR-4 device at power-up. After the rising edge of $\overline{\text{RESET}}$, Φ_{LC} is obtained from the hardware oscillator (OSCCON<SCO> = 0) with a 1:32 prescaling ratio (OSCCON<PS2:PS1> = 011b = 3). If a standard 32 kHz watch crystal is used, Φ_{LC} oscillates at 1 kHz, a suitable inversion frequency for most nematic liquid crystal modulators.

NBIAS is a bias voltage for n-channel current sources in the operational amplifiers and comparators of DHR-4. Its value is not critical; 1.5 V works well.

3.2.2 Pixel circuit

At the core of each pixel is a static memory bit implemented using cross-coupled CMOS inverters (see Figure 3.5). This 1-bit memory can latch data presented to the pixel optically, via a thresholding detector and its associated buffers (top left in Figure 3.5), or electronically, through a microprocessor interface bus connected to the two lines on the lower right corner of Figure 3.5. The state of the latch can also be read using the bi-directional microprocessor interface bus. An analog multiplexer (lower left corner of Figure 3.5) drives the liquid crystal electrode based on the state of the latch.

[4]This oscillator can operate with either a crystal (OSCMODE = 0) or a simple RC network (OSCMODE = 1).

[5]See Figure 3.2.

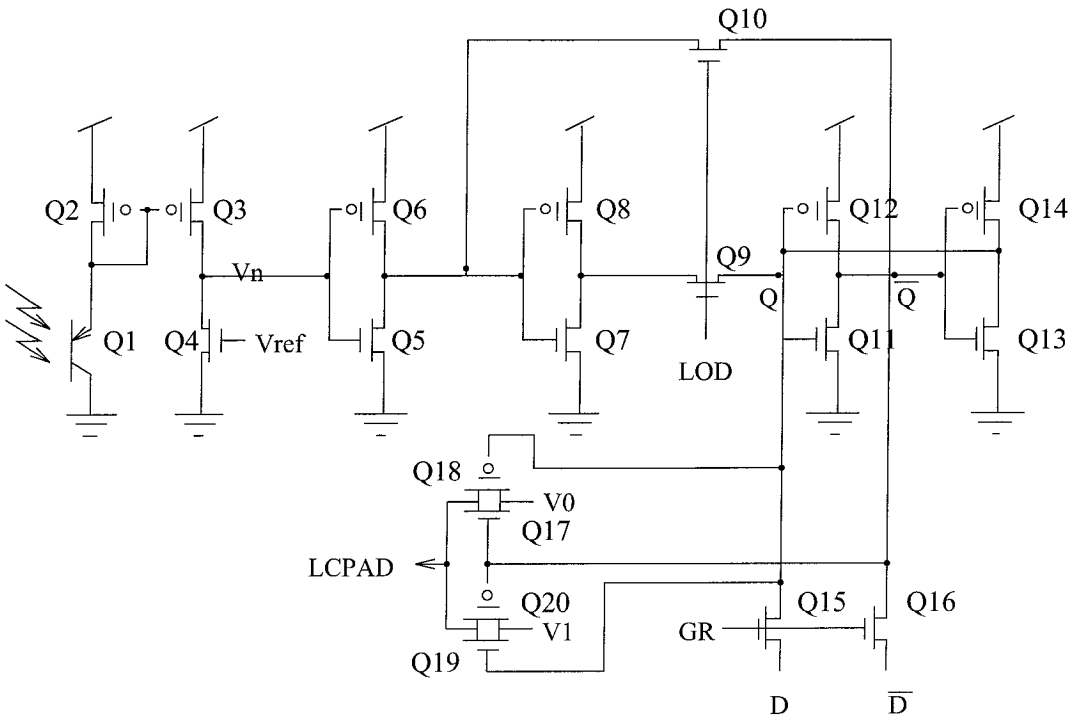


Figure 3.5: Schematic of a dynamic hologram refresher pixel.

Q1 is an active-well-substrate PNP bipolar junction phototransistor with a floating base. Q2 and Q3 form a current mirror that isolates the threshold computation from the detection process. Q2–Q4 are normally operated in the weak inversion (subthreshold) regime, where the saturated drain current is an exponential function of the gate-source voltage[28, 98]:

$$I_d = \frac{qAD_n n_i e^{-q\psi_B/(kT)}}{L} e^{q\psi_s/(kT)} \left(1 - e^{-qV_d/(kT)}\right) \tag{3.6}$$

$$= I_0 \frac{W}{L} e^{q\kappa V_g/(kT)} \left(1 - e^{-qV_d/(kT)}\right) \tag{3.7}$$

(for an n-channel MOSFET with a grounded source and negligible Early effect) where q is the electron charge, D_n is the electron diffusivity, n_i is the intrinsic carrier concentration, k is Boltzmann's constant, T is the temperature, ψ_s is the surface potential, V_d is the drain voltage, V_g is the gate voltage, A is the effective transverse area of the region where current flows, and W and L are respectively the width and length of

the MOSFET. The surface potential ψ_s is approximately a linear function of the gate voltage:

$$\psi_s = \psi_{s0} + \kappa V_g + o(V_g). \tag{3.8}$$

Because of the exponential dependence of the current flowing through Q1 and Q2 on the emitter voltage of Q1, this voltage does not vary by more than a few hundred millivolts over the entire optical power input range, and the collector-base junction of Q1 is always strongly reverse biased. This ensures that the collector-base depletion region is thick and that the electric field in it is high[27]. In turn, this allows more photogenerated carriers to be captured and strongly accelerated in the depletion region.

Q3 and Q4 form a programmable thresholding circuit with current input and voltage output. The shape of the threshold function is determined by the geometry of Q3 and Q4 instead of the properties of Q1, which are more difficult to control. The level of the threshold is set by V_{ref}. The circuit compares the photocurrent flowing through Q3 to Q4's saturation current. Figure 3.6 shows the drain characteristics of Q3 and Q4. The same current I must flow through both devices. Therefore, The output

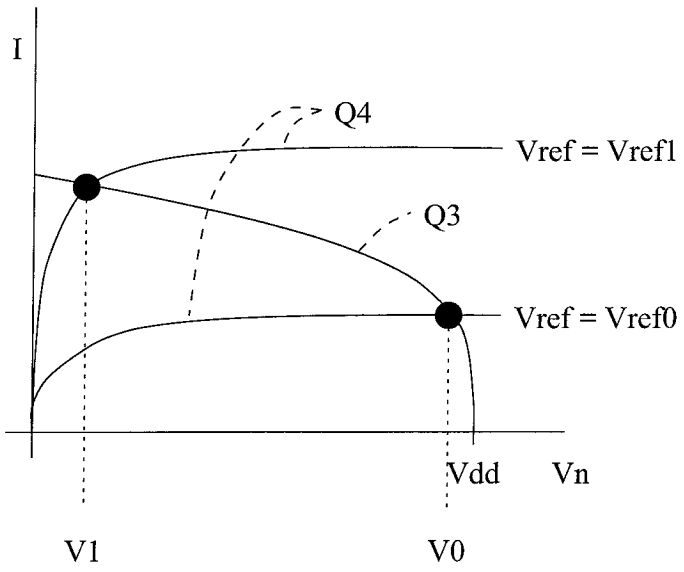


Figure 3.6: Operation of the thresholding circuit.

voltage V_n is given by the intersection of the drain characteristics of the two devices. Let us consider the case where the photocurrent is such that the drain characteristic of Q3 is the curve labeled Q3 in Figure 3.6. At a low V_{ref} setting, $V_{\text{ref}} = V_{\text{ref0}}$. The curves intersect near the upper rail, and the output voltage is V_0. As V_{ref} is increased, the current flowing through Q4 at any output voltage increases, while the current through Q3 remains constant. The output voltage "glides" along the drain characteristics of the two transistors. At $V_{\text{ref}} = V_{\text{ref1}}$, the output voltage is close to zero: $V_n = V_1$. The thresholding curve (V_n vs. optical power) can be made sharper by reducing the drain conductance of either or both Q3 and Q4, e.g., by elongating their channels to suppress the Early effect.

Q5–Q8 are two CMOS inverters that sharpen the threshold characteristic of the circuit and avoid loading the analog front end when optical data are latched. Q9 and Q10 are NMOS transmission gates activated by LOD (*Load Optical Data*). When LOD is asserted, the complementary outputs of the two inverters to the left of Q9 and Q10 drive the regenerative latch (Q11–Q14). Q15 and Q16 are pass transistors that connect the memory bit to the differential column data bus (D and $\overline{\text{D}}$). All accesses to the memory circuit are differential.

3.2.3 Implementation

A DHR die (4.6 mm×6.8 mm) fabricated through MOSIS[78] in Orbit Semiconductor's inexpensive 2.0 μm double-metal n-well process is shown in Figure 3.7. The shiny first-level-metal liquid-crystal electrodes associated with each pixel are clearly visible. To the left and bottom of the pixel array are row and column address decoders, respectively. The oscillator, analog multiplexers and two special-function registers are visible above the array, while the microprocessor interface circuitry lies below the array. Figure 3.8 shows the pin-out of this device.

The size of the pixels (including interconnection and power distribution overhead) is 132 μm × 211 μm. The layout of the pixel circuit is shown in Figure 3.9. Virtually all the optically inactive area of the pixel is shielding from light, mostly by second-level

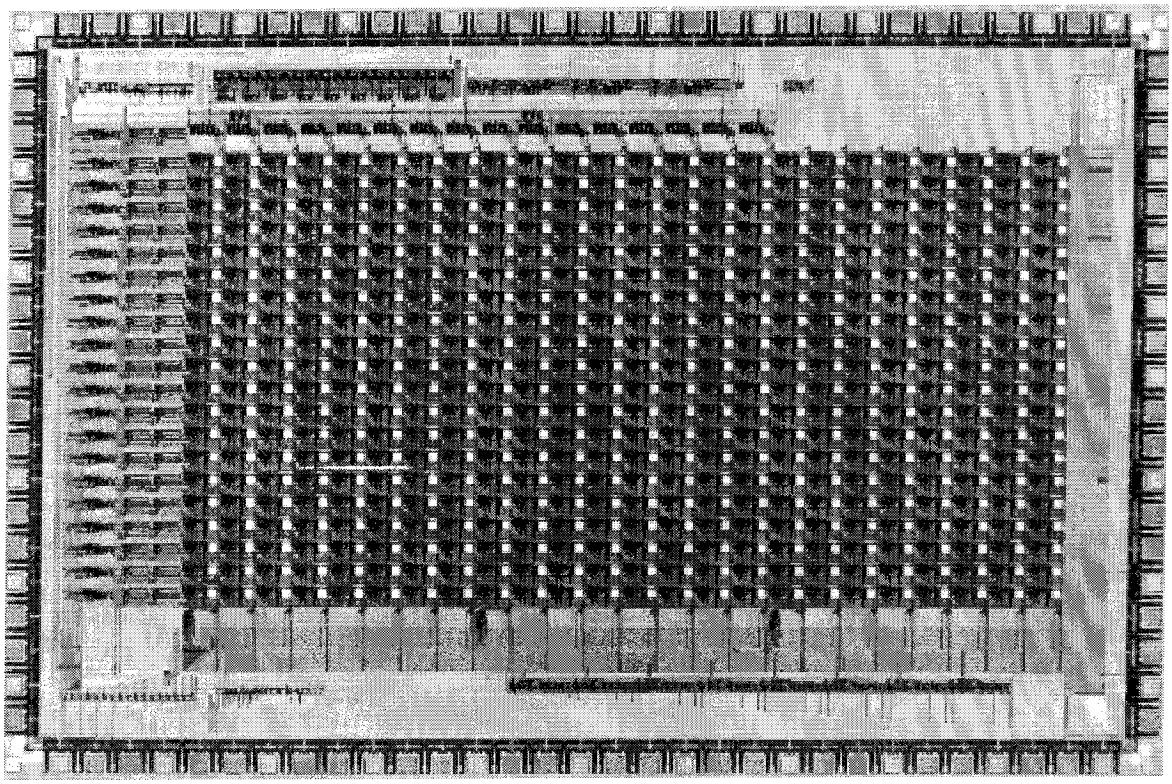


Figure 3.7: Photograph of a DHR-4 die.

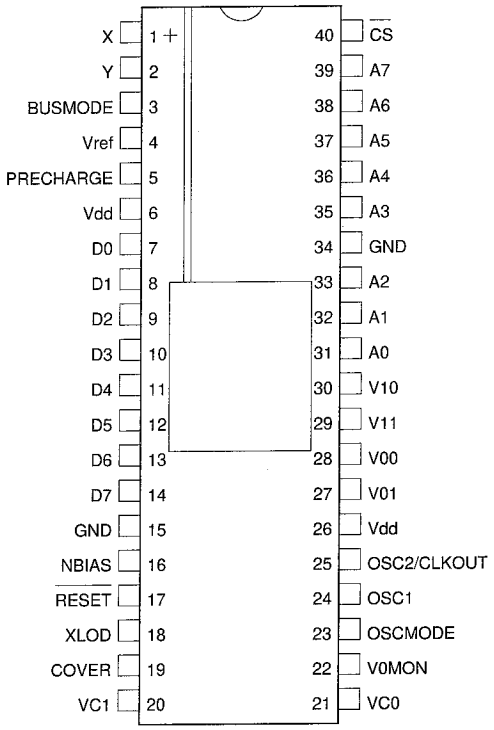


Figure 3.8: Pin-out of the DHR-4 OEIC.

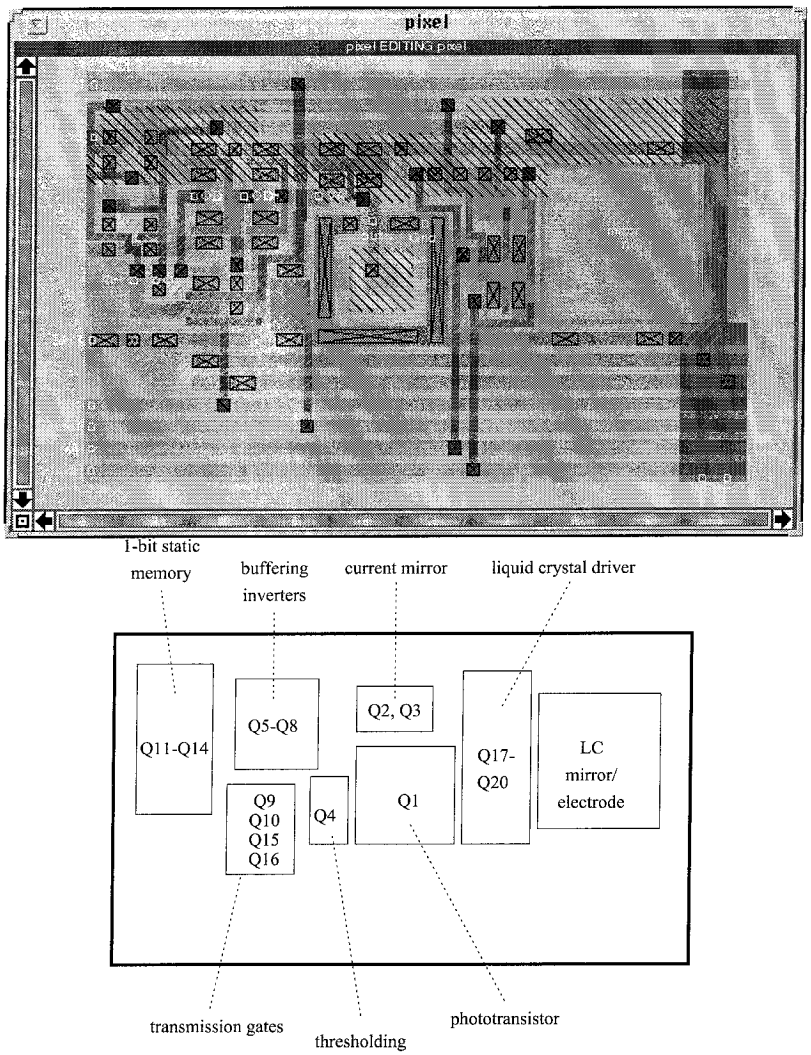


Figure 3.9: Layout of the DHR-4 pixel circuit.

metal, but also, in the rightmost part of the pixel, by first-level metal and polysilicon. A large 49 μm $\times$ 49 μm first-level metal electrode drives the overlying liquid crystal layer and serves as an optical mirror. Ideally, these functions would be performed by the second-level-metal light shield overlying most of the pixel's circuitry; the modulator electrode would then require no dedicated area. However, the optical quality of the second-level metal provided by the process in which the device was fabricated was poor, so first-level metal was used.

The dimensions of the MOSFETs are listed in Table 3.2 (see also Figures 3.5 and

MOSFETs	Dimensions (width/length)	Function
Q2, Q3	8 μm/4 μm	current mirror
Q4	8 μm/16 μm	thresholding (current source)
Q5, Q6, Q7, Q8	8 μm/2 μm	buffering inverters
Q9, Q10, Q15, Q16	8 μm/2 μm	pass transistors
Q11, Q12, Q13, Q14	4 μm/2 μm	regenerative latch
Q17, Q18, Q19, Q20	8 μm/2 μm	analog multiplexer

Table 3.2: Dimensions of the MOSFETs in the DHR-4 pixel circuit.

3.9). Q4 has a long channel to minimize its drain conductance (Early effect) and sharpen the threshold function.

The phototransistor is a parasitic active-well-substrate bipolar junction phototransistor. Its emitter consists of a 10 μm $\times$ 10 μm (drawn dimensions) p-type active area in a 20 μm $\times$ 20 μm n-doped well, centered on the 18 μm $\times$ 18 μm opening of the overlying second-level metal light shield (see Figure 3.9). Part of the opening is obstructed by the emitter connection, which consists of a 4 μm $\times$ 4 μm emitter contact in the upper left corner of the emitter and a 3 μm $\times$ 4 μm segment of first-level metal wire.

HAN-on-VLSI modulators were fabricated on DHR die using the procedure described in Chapter 2. The 5 μm cell gap was filled with Merck E7[90].

We are confident that much denser arrays could easily be obtained by employing a more advanced process with a smaller minimum feature size and a topmost metallization level featuring a better optical quality.

This device can be used in a variety of optical memory and neuromorphic systems, where it is employed as an electrically addressed spatial light modulator (SLM) to record holograms, as a thresholding detector array for readout, and as a latching optically addressed SLM to refresh dynamic holograms that slowly decay as a result of successive accesses to the holographic recording material. In this last function, the DHR detects, thresholds and latches a hologram being read out, and displays the latched information on its modulators to refresh the hologram.

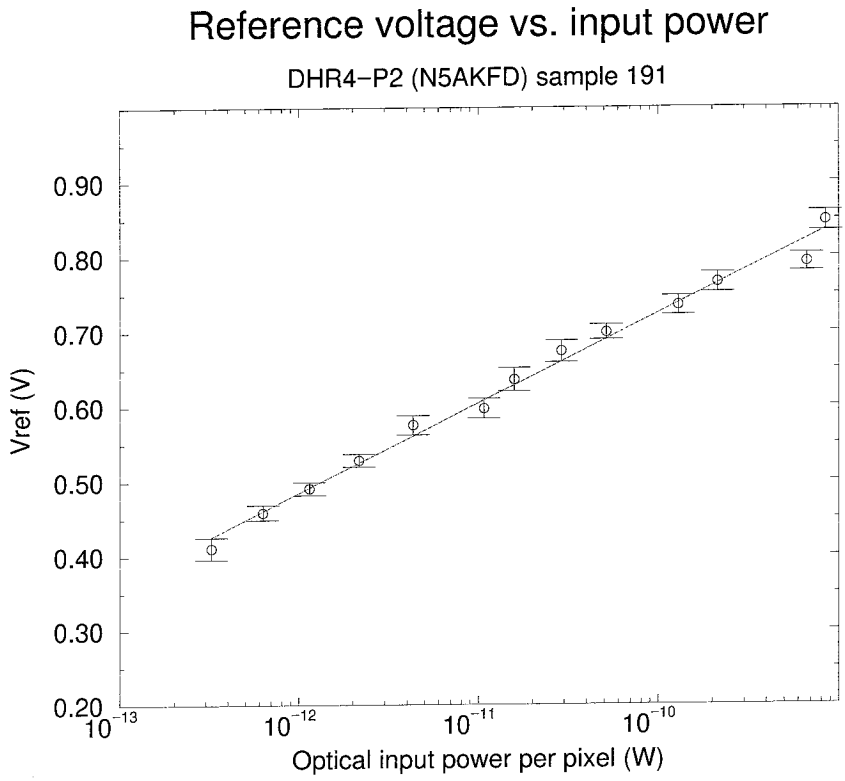


Figure 3.10: Input-output characteristic of the optical receiver (detector and thresholding circuit).

3.2.4 Experimental results

Optical detection

The DHR-4 OEIC does not allow the independent and direct characterization of its photodetectors and thresholding circuits, primarily because of pin-count limitations. However, the global input-output characteristic of these components, i.e., V_{ref} vs. incident light power, can be measured. The OEIC was illuminated with a uniform truncated plane wave at 488 nm. For each of 13 different intensity levels, V_{ref} was scanned, and the value of V_{ref} at which each pixel changed states was recorded. The sample mean and standard deviation of this V_{ref} value was computed over a portion of the pixel array containing 16×21 pixels. The results are shown in Figure 3.10. The height of the error bars is twice the sample standard deviation of the measured V_{ref}.

Since the current flowing through Q4 (see Figure 3.5) is proportional to the optical input power, the surface potential linearization parameter, κ, can be extracted from Figure 3.10. The slope of the optical receiver's input-output characteristic is

$$\frac{kT}{q\kappa}.$$

A logarithmic fit (solid line in Figure 3.10) gives $\kappa = 0.5$, assuming $T = 295$ K.

In a separate experiment, we tried to determine the minimum illumination level that would be sufficient for DHR-4 to correctly detect and memorize a data page. Optical patterns were latched in the on-chip memory with less than 100 fW incident on each phototransistor.

While the DHR-4 OEIC does not allow a direct measurement of the response of its phototransistors, another OEIC designed by the author employs identical photodetectors, was fabricated in the same process (albeit in a different run), and permits a measurement of the device's input-output characteristic. Figure 3.11 shows the responsivity curve of a photodetection circuit on this OEIC at $\lambda = 632$ nm. The circuit consists of a PNP phototransistor identical to the one shown in Figure 3.5 and of a 20:1 PMOS current mirror. Both parts of the circuit were shielded. They were also located more than 4.5 mm away from the bonding pad carrying the output current, in order to eliminate any optical response from the electrostatic-discharge protection diodes in the pad cell. The responsivity of the phototransistor[99],

$$(1 + h_{\mathrm{FE}})\,\frac{q\eta\lambda}{hc}, \tag{3.9}$$

where η is the detector's quantum efficiency, h is Planck's constant and c is the speed of light in vacuum, was found to be

$$R_{\mathrm{PNP}} = 11.5 \text{ A/W}. \tag{3.10}$$

If we assume that[25, Figure 11-15] $\eta \approx 0.3$, we obtain the following estimate for the

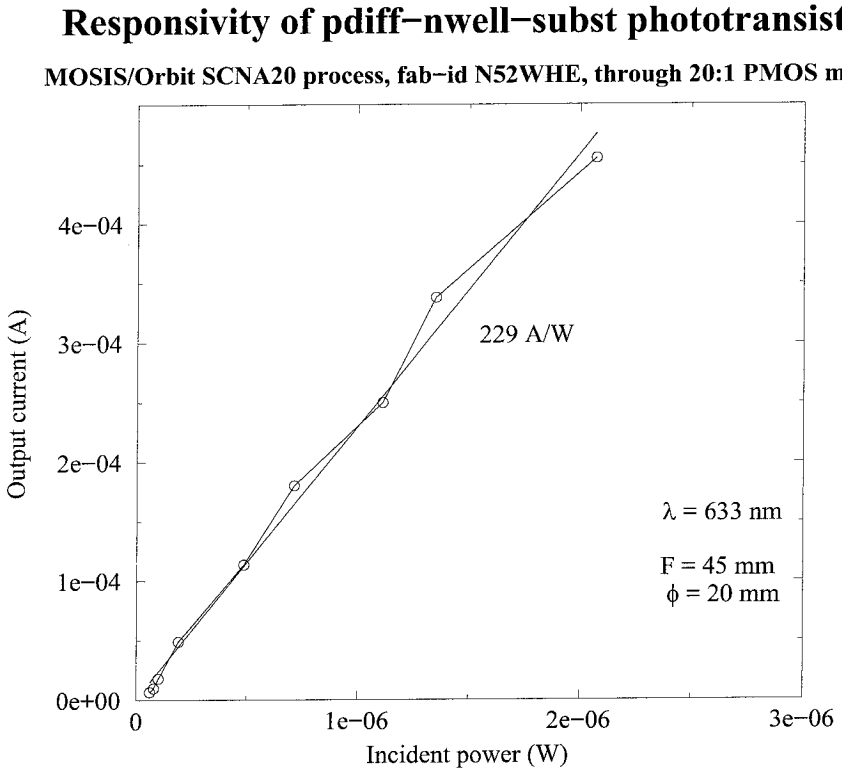


Figure 3.11: Input-output characteristic of a PNP phototransistor at 632 nm.

DC forward current transfer ratio of the phototransistor:

$$h_{\mathrm{FE}} \approx 74. \tag{3.11}$$

We expect the above calculations to somewhat underestimate R_{PNP} and h_{FE}, because

- the focused spot may have been slightly bigger than the opening in the second-level metal shield surrounding the active area of the photodetector, resulting in a loss of optical power;
- the shadowing effect of the wire connecting to the emitter was not taken into account.

The responsivity is expected to drop at shorter wavelengths, e.g., at the blue-green wavelength ($\lambda = 488$ nm) used in our holographic storage experiments. The responsivity decrease is due to the following effects.

- The responsivity is proportional to the wavelength (see equation 3.9).
- The absorption increases when the wavelength decreases. At $\lambda = 488$ nm, the absorption length is less than 1 μm[100]. Fewer photons reach the immediate vicinity of the collector-base junction at a depth of approximately 3 μm.

In the end, despite our interest in characterizing the photodetectors, what matters the most from a system-level perspective is the global input-output characteristic of the complete receiver, consisting of a phototransistor, a current mirror and a simple thresholding circuit. This is the curve shown in Figure 3.10. The uniformity of the modulators (related to the error bars in Figure 3.10) was sufficient to correctly detect data pages in our experiments, because of the good contrast ratio of the modulators (see §3.2.4). However, should an improved optical-receiver uniformity be desired, hot-electron injection could be employed to add an offset to the surface potential of Q3, for example[101]. This form of circuit trimming requires no additional devices in the pixel circuit (see Figure 3.5).

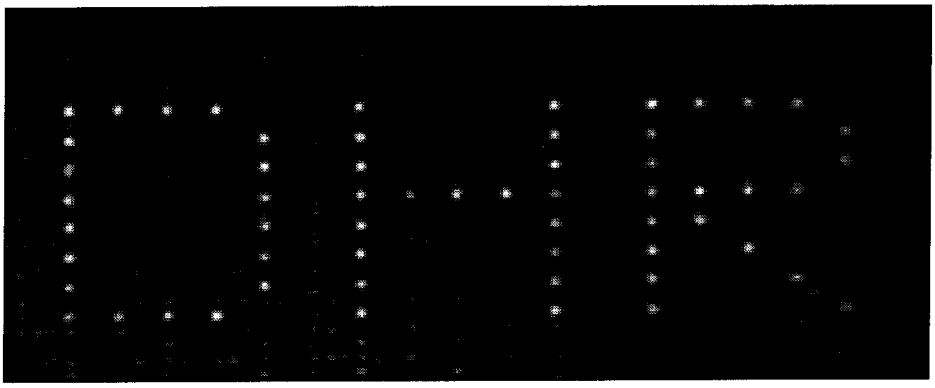


Figure 3.12: A DHR-4 display.

Optical modulation

While some of the earlier DHR-4 devices exhibited some relatively serious modulator non-uniformities (see, e.g., the photographs of sample 175 in Figure 6.4), the electro-optic response of more recent samples is considerably improved. The improvement is due in part to our group's new clean room in the Moore Laboratory.

Figure 3.12 shows DHR-4 sample 191[6] displaying its name using its HAN-on-VLSI modulators. The device was illuminated with a collimated plane wave at 488 nm.

We characterized the properties of the subset of the modulator array used to display binary patterns in Figure 3.12 and in the experiments of §6.3. The liquid-crystal driving voltages were optimized for contrast over the entire array. Frames were then captured, with all pixels on and all pixels off. Statistics of the on and off pixels were computed, as were the contrast ratios of individual modulators. Only the globally optimized liquid-crystal voltages were used; the voltages were not optimized for individual pixels.

Figure 3.13 shows histograms of the reflectance of the modulators in the on and off states. The reflectance is represented in arbitrary units in Figure 3.13. More accurately, the reflectance is coded as the digital output of a camera and video digitizer combination. The detection linearity of the setup was verified before the experiment was performed. Note that the distributions do not overlap; there are separated by

[6]This is the same sample used in the experiments discussed in §6.3.

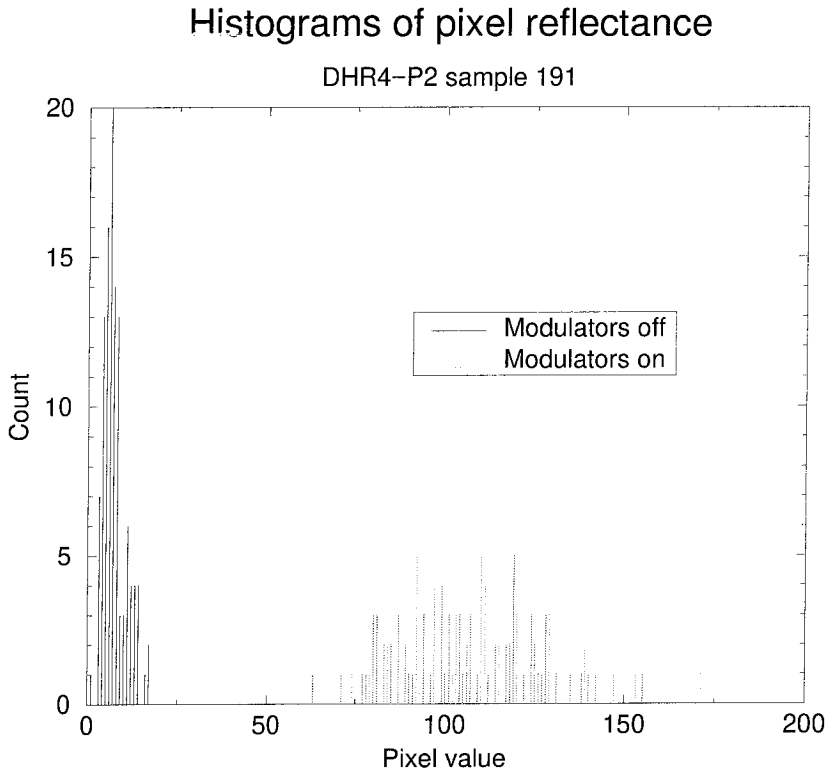


Figure 3.13: Histograms of the modulators' reflectance in the on and off states.

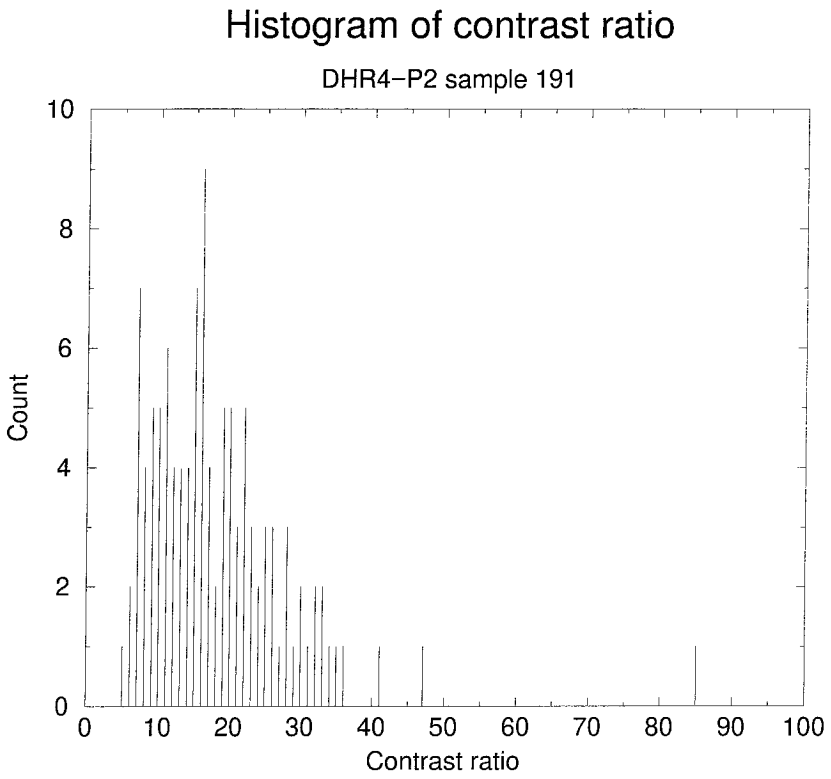


Figure 3.14: Histogram of the modulators' contrast ratios.

46 gray levels. The mean gray-level codes of the zero and one (off and on) distributions are respectively 7.2 and 107; their sample standard deviations are 3.4 and 19.8, respectively.

The contrast ratio[7] of each modulator was measured. The distribution of the contrast ratios is presented in Figure 3.14. The mean contrast ratio is 18.3:1; the sample standard deviation is 10.5.

3.2.5 Array density

The main drawback of the DHR-4 pixel design is its size, approximately 28000 μm^2, which limits the number of pixels that can be incorporated in a reasonably sized die. The large size of the pixel is due primarily to the following factors.

[7]The contrast ratio is defined as the ratio of the reflectance of the modulator in the on state to its reflectance in the off state.

Feature size. DHR-4 was implemented in an inexpensive 2.0 μm process, while 0.35 μm processes are now commonly used to fabricate microprocessors and memory integrated circuits. Assuming that the design rules are scalable (and this is not a very restrictive assumption), the pixel area is approximately proportional to the square of the minimum feature size.

Number of metallization levels. The process employed to fabricate DHR-4 has only two metallization levels, one of which was used as a light shield and covered most of the area of the pixel. This left only one level of metallization for general purpose interconnections and power distribution. The resulting layout has large areas employed only by first-level metal interconnections and a second-level metal light shield.

Optical quality of topmost metal level. Ideally, the topmost metallization level could be employed as an electrode and as a mirror for the reflective optical modulators, in addition to its function as a light shield. However, the poor optical quality of the second-level metal in the process employed to fabricate DHR-4 called for an area of first-level metal to be set aside in each pixel for use as a modulator electrode and mirror (see Figure 3.9). In order to ensure the flatness of the resulting mirror, no circuits were drawn under it. This area is therefore electronically wasted.

All the difficulties enumerated above can be overcome by employing a more advanced process. For example, if a 3-metal process with highly reflective and planar third-level metal is assumed to be available, and if this process employs design rules similar to the MOSIS scalable CMOS (SCMOS) rules[78], it is estimated that an updated DHR-4 pixel could be drawn in only 9000 λ^2.[8] If it is fabricated using a 0.35 μm process, this pixel could measure approximately 16.6 μm $\times$ 16.6 μm, or less than 1% of the area of the original DHR-4 pixel.

[8]The resolution of a VLSI layout is customarily called λ. In the MOSIS scalable CMOS rules, λ is one half of the minimum feature size (the minimum MOSFET gate length). For example, in the Orbit 2.0 μm process, $\lambda = 1.0$ μm.

3.3 Dynamic DHR pixels

The savings in pixel area achievable by employing a 3-metal process with a small feature size are impressive and useful. However, one wonders if an even higher pixel density could be obtained by using a more aggressive pixel design. One important motivation for increasing the number of pixels per data page is the fact that the pixel diffraction efficiency of the holograms (the ratio of the power diffracted into one pixel area to the reference wave power) goes as

$$\eta_{\text{pixel}} \propto \frac{1}{M^2 N},$$

where M is the number of holograms and N is the number of pixels per page. Hence, given a fixed light budget (i.e., a fixed minimum diffraction efficiency, determined by the optical input power and the sensitivity of the photodetectors), the maximum storage capacity MN is achieved by storing as many pixels as possible within each hologram.

Small additional savings could perhaps be obtained by eliminating some of the buffering and isolation transistors (e.g., Q2, Q3, Q5 and Q6) and by employing only NMOS transistors in the transmission gates that drive the liquid crystal modulator (at the cost of a less effective DC balancing and hence a potentially shorter modulator lifetime). However, extensive additional savings can be achieved by foregoing the convenience of employing a static memory and by sharing some pixel components between all three functions of the pixel: optical detection, memory, and optical modulation (see Figure 3.1). The number of MOSFETs can be decreased to four or even two, depending on the available optical power[102].

3.3.1 Two-transistor pixel

Figure 3.15 shows the most compact proposed pixel circuit, consisting of a photodiode D, a capacitor C and two NMOS transistors, Q1 and Q2. The liquid crystal electrode is connected to the common node of Q1, Q2 and C.

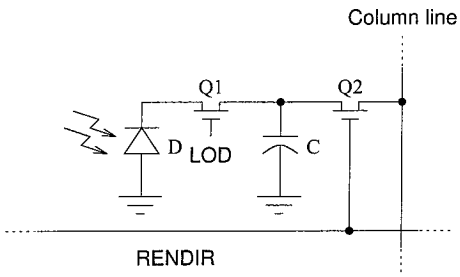


Figure 3.15: Two-transistor DHR pixel. The liquid crystal electrode is connected to the common node of Q1, Q2 and C.

First consider Q2 and C. Together, they form a conventional dynamic RAM cell. The capacitor C stores a charge that represents one bit of data. All the cells in a column of the cell array share a column line, employed to read, write and refresh data stored in the cells. All the cells in a row share a row addressing circuit that generates a row enable signal, RENDIR. At most one row is selected at any given time. Q2 connects C to the column line when RENDIR is asserted. The data stored on the capacitor can be read by monitoring the change in the voltage of the column line using a sensitive amplifier. New data is written in the cell by driving the column line (and the pixel's capacitor) to the desired voltage level. The charge stored on the pixel capacitors must periodically be refreshed, because it slowly decays due to leakage currents and parasitic conductances. This is accomplished by reading the data stored in an entire row into a 1-D temporary storage array at the bottom of the main array, and rewriting the data into the row. The refresh period of modern dynamic memories typically ranges from 32 ms to 128 ms. The storage capacitance C is ordinarily a few tens of femtofarads.

The same cell architecture, consisting of one MOSFET and one capacitor (Q2 and C), is also the pixel circuit of a simple active-matrix dynamic liquid crystal display. Each pixel stores the voltage needed to drive one pixel. The rows are addressed in succession. The columns are driven with the voltages to be assigned to an entire row of pixels. The RENDIR signal of that row is asserted (all the other rows have RENDIR = 0) to allow the capacitors to "memorize" the desired voltages. RENDIR

is then deasserted, and the process is repeated for another row. In a spatial light modulator, the rows must be addressed periodically not only because of the decay of the stored charges, but also because liquid crystal modulators must be driven with DC-balanced signals[9] at a frequency typically on the order of 1 kHz. The transistors and capacitors can be fabricated on transparent substrates such as glass coated with polycrystalline or amorphous silicon (this is the case of most advanced transmissive liquid crystal displays using thin-film transistors), or on a silicon die in liquid-crystal-on-silicon (LCOS) modulators.

The new pixel circuit uses Q2 and C as an active-matrix display would use these elements when data must be displayed (i.e. when recording or refreshing a hologram). It uses the same components much as a dynamic RAM during readout. The capacitor C is also employed to detect optical signals. Optical detection proceeds in two steps. First, all the capacitors are charged to a known voltage. Then, LOD is asserted, turning Q1 on, while the RENDIR signals of all the rows are close to 0 V. In each pixel, the charge stored on the capacitor is discharged through the photodiode D, which acts as a current source controlled by the optical power incident on the pixel. More specifically, the photocurrent flowing through the diode (provided by the capacitor) is

$$I_{\text{ph}} = \frac{Pq\eta}{h\nu}, \tag{3.12}$$

where P is the incident power, q is the electron charge, η is the quantum efficiency of the detector, h is Planck's constant, and ν is the frequency of the light. The voltage change on the capacitor is

$$\Delta V = \frac{I_{\text{ph}} t_{\text{int}}}{C}, \tag{3.13}$$

where t_{int} is the integration time. Equations 3.12 and 3.13 are valid if I_{ph} is considerably larger than currents due to leakage and parasitic conductances, and the voltage stored on the capacitor at the end of the integration period is still significantly larger than the "thermal voltage" kT/q (25 mV at room temperature), where k and T are

[9] DC-balancing the driving signals helps in obtaining a long display lifetime. It is believed that selective ion trapping happens in the liquid crystal alignment layers if the polarity of the signals is not inverted frequently enough.

respectively Boltzmann's constant and the temperature of the photodiode. At the end of the integration time, LOD is deasserted, and the rows are successively read using the mechanism described above for a dynamic RAM. If the holograms needs to be refreshed, the logic levels read from the pixel array are converted into appropriate liquid crystal voltages and written to the array, one row at a time.

We have seen that the capacitor C is shared between the three functions of the pixel: optical detection, optical modulation and memory. Only two MOSFETs and one photodetector must be added to form a complete dynamic DHR pixel. In comparison, the static DHR-4 pixel (shown in Figure 3.5) has 19 MOSFETs and one phototransistor. Notice also that the two transistors are of the same type (native or n-channel) and that there are no connections to the upper supply rail V_{dd}. Therefore, there is no need to diffuse n wells in the p substrate (assuming an n-well process) or to devote area to well plugs. Wells are expensive in terms of silicon real estate, because native and well transistors must be separated from their boundaries by a sizable distance. Moreover, only one supply rail (ground) must be distributed to the pixel array.

The price to pay for this drastic reduction of the number of components in the pixel circuit is a more complex on-chip support circuitry. For example, when a data page must be latched in DHR-4's internal memory, it suffices to pulse LOD high. The data will remain in the regenerative latch (Q11–Q14 in Figure 3.5) as long as the device is powered, and can be read out at the convenience of the circuit that requests the read operation. In the case of a DHR based on the new compact pixel architecture, the pixel capacitors must first be precharged. Then, LOD must be pulsed high for a carefully controlled integration time t_{int}. Finally, the data must be read out promptly from the pixels, before the capacitors discharge. The pixel readout process is also itself slightly more complicated than in the case of DHR-4; the latter provides complementary signals with rail-to-rail voltage swings, while the former uses single-ended column lines in the interest of compactness. The increased complexity, however, is not prohibitively expensive. The area of the control functions grows roughly as the square root of the number of pixels. Its development cost would be largely offset by

the significantly improved pixel density provided by the new design.

3.3.2 Four-transistor pixel

The main difficulty currently anticipated with the pixel circuit shown in Figure 3.15 is the small voltage swing (equation 3.13) produced by low optical input powers. This problem can be alleviated to some extent by sacrificing speed (increasing t_{int}) for sensitivity, but the integration time must remain significantly less than the dark decay time of the charge stored on C. Small voltage changes, corresponding to minute amounts of charge stored on the small capacitor C, are difficult to detect using a DRAM-type pixel such as the one shown in Figure 3.15 because the detection mechanism involves redistributing the charge stored on C between C and a much larger column line capacitance. The capacitance to ground of the pixel storage node is small (in a dynamic RAM, it is only of the order of 15 fF to 50 fF, to minimize the silicon area devoted to the capacitor), whereas the column line has a large capacitance C_{line}, consisting of contributions from the metal line itself, from source/drain diffusions and from the gates of the CMOS circuits connected to the line. When a pixel is read, the charge it stores is redistributed between the pixel capacitance and the column line capacitance; the effect of this redistribution is that the optically generated voltage swing is multiplied by

$$\frac{C}{C + C_{\text{line}}},$$

a ratio much smaller than one.

Figure 3.16 shows a modified compact pixel that addresses this problem. This pixel is identical to that shown in Figure 3.15, except for the addition of two n-channel MOSFETs, Q3 and Q4. The operation of the circuit is as described in §3.3.1, except when reading out optical signals.

When a data page must be read out, the pixel capacitor is precharged to the voltage of the column line by pulsing RENDIR high and the photocurrent is integrated on C during the LOD pulse. However, the resulting charge is not sensed by sharing it

with the column line; instead, a source follower[10] buffers the voltage and drives the column line. The input transistor of the source follower is Q3. A single bias transistor is shared by all the pixels in a column of the array. At most one input transistor is connected to (put in series with) the bias transistor at any time: transistor Q3 of the pixel at the intersection of the column and the selected row (RENFOL=1). The row addressing circuit of the selected row generates an active (high) RENFOL. This turns on Q4, and connects Q3 to the source follower bias transistor of the column to which the pixel belongs. RENFOL=0 in all the other rows. The output voltages of the selected row of pixels appear at the drains of the source follower bias transistors (one per column). Note that the charge stored in the pixel is not shared with the column line. It is instead buffered by a source follower, which drives the column line by charging or discharging it with its saturation current. This circuit is susceptible to errors due to threshold variations of Q3. However, this can be compensated for using correlated double sampling[103] (the precharge voltage, as buffered by the source follower, is subtracted from the optically generated signal). The incorporation of the input transistor of a source follower (Q3) and a switch activating it (Q4) were inspired by a similar technique used in active-pixel image sensors[36].

The circuit shown in Figure 3.16 can perform all the functions of the original DHR, with only 4 MOSFETs, one photodiode and one capacitor. Note that all the transistors are still of the same type; there is therefore no need for an n-doped well. Moreover, area can be saved by laying out the source of Q1 as an extension of the cathode of the photodiode (assuming an active-substrate photodiode is employed). Likewise, the source of Q3 and the drain of Q4 can be a shared area of active, with the gates of the two transistors separated by the minimum poly-poly separation (2 λ in the MOSIS scalable CMOS rules). This circuit is estimated to require only 64 μm^2 (or 0.2% of

[10]A source follower is a voltage amplifier with a gain close to unity. In its simplest form, it consists of two transistors of the same type (in our case, native or n-type MOSFETs) in series. One transistor, called the bias transistor, has its source grounded and its gate held at the bias voltage, V_{SFNBIAS}. Its drain, which is connected to the source of the second transistor, is the output of the amplifier. The second transistor has its drain connected to the upper supply rail V_{dd}, and receives the input voltage on its gate. Assuming both transistors operate in the saturation regime, the output follows the input.

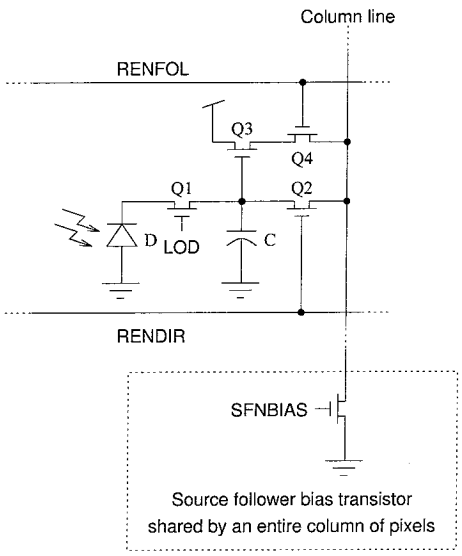


Figure 3.16: Four-transistor DHR pixel.

the area of the original DHR-4 pixel) if it is fabricated in a suitable 0.35 μm DRAM process.

Finally, under certain circumstances, it is possible to remove transistor Q1 from the circuits illustrated in Figures 3.15 and 3.16. Q1 serves two main functions:

- it prevents the premature decay of the voltage stored on C during display (recording or refresh) cycles;
- it helps provide a well-controlled integration time during readout cycles.

Q1 can be eliminated if the following two conditions are met:

- the detectors are not strongly illuminated during display (recording or refresh) cycles (e.g., an array illuminator focuses illumination beams on the modulators);
- the integration time during readout cycles is accurately controlled through RENDIR.

3.3.3 Noise analysis

An experimental characterization of DHR-4's optical receivers showed that the device can reliably detect and latch binary patterns with less than 100 fW incident on each

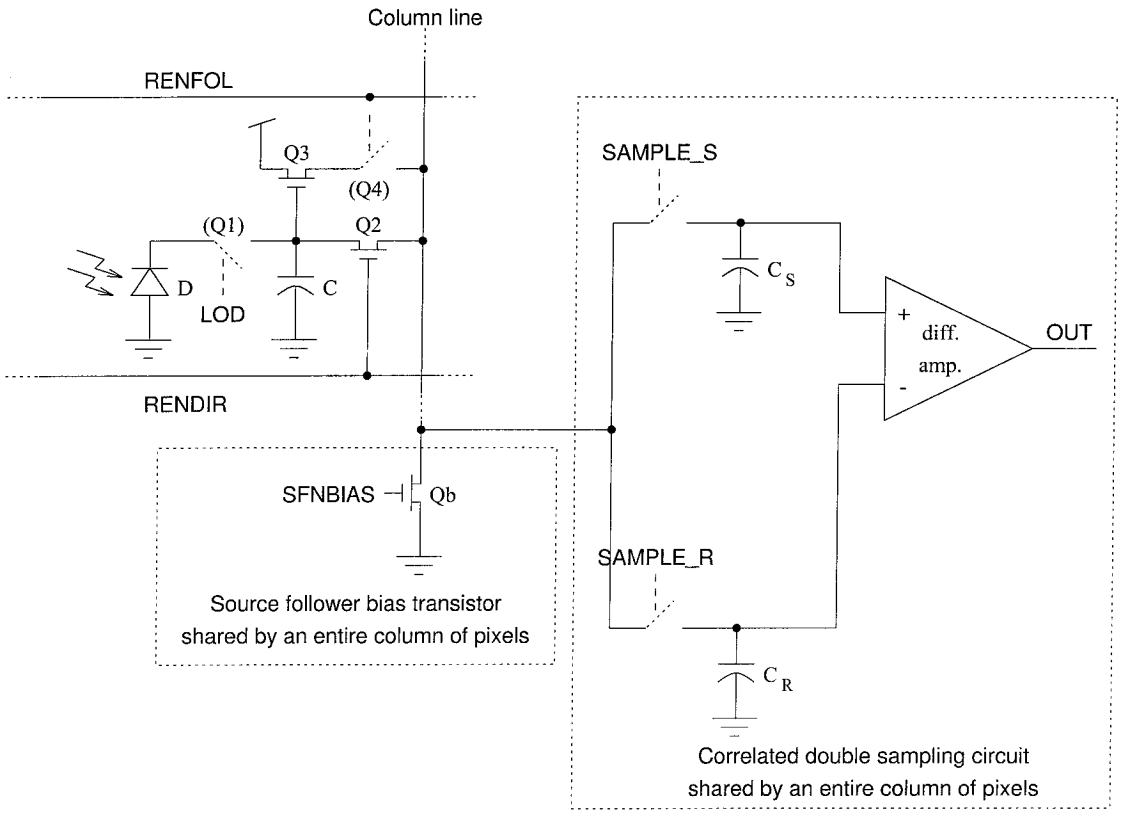


Figure 3.17: Circuit used for the noise analysis of dynamic DHR pixels.

"on" pixel. While the dynamic DHR pixels described above have not yet been fabricated, we can estimate their sensitivity by performing a noise analysis. Figure 3.17 shows the simplified circuit used in this analysis. Q1 and Q4 are modeled as ideal switches (alternatively, Q1 could be removed altogether, as explained above). Correlated double sampling[103, 36] is employed to compensate for offsets in the input transistors of the source followers (Q3 and Qb) and to suppress the pixels' kTC noise and the source followers' flicker ($1/f$) noise. The reset voltage on capacitor C is sensed and stored (with a fixed offset due to the source follower) on capacitor C_R at the end of the precharge that precedes the integration of an optical signal. At the end of this integration, the remaining voltage on C (the signal voltage) is read and stored on C_S (with the same offset caused by the source follower). A difference amplifier computes the difference between the voltages stored on C_S and C_R, which is proportional to the

total exposure of photodiode D during the integration time t_{int}. Any non-uniformities due to threshold voltage variations between input transistors (Q3) of different pixels are suppressed. Such non-uniformities appear as a constant voltage added to both the signal (S) and reset (R) levels, and are eliminated by the difference amplifiers; there is thus no need for intra-pixel offset cancellation. The uncertainty (Nyquist noise) in the reset voltage stored on C is also removed by correlated double sampling, because it is present (and correlated) in both the signal and reset levels read out by the source follower. Moreover, correlated double sampling inherently removes low-frequency components of the output of the source follower. Therefore, flicker noise is suppressed.

We can identify the following independent sources of noise in the circuit of Figure 3.17.

Signal shot noise This is the noise due to the quantization of the photogenerated charge carriers, or, equivalently, of the particular nature of the input light signal.

Dark-current shot noise A reverse current due to thermally generated charge carriers flows through the reverse-biased photodiode D. There is also a noise component associated with this current.

Nyquist noise of storage capacitor The precharge operation causes an uncertainty in the amount of charge stored on C.

Nyquist noise of source follower This is the noise caused by the thermal velocity and lifetime distribution of charge carriers in Q3 and Qb.

Flicker noise of source follower This is the $1/f$ noise (due to traps) in transistors Q3 and Qb.

Nyquist noise of sample-and-hold circuit This is the kTC noise present on capacitors C_S and C_R.

Nyquist and flicker noise due to difference amplifier The thermal and $1/f$ noise components due to the difference amplifier.

We will now discuss each noise contribution individually, and explain why certain components can be disregarded. We will formulate an estimate of the total noise referred to signal storage node (capacitor C).

Signal and dark-current shot noise

The mean-square current due to signal and dark-current shot noise is simply[103, 100, 25]

$$\overline{Q^2_{\text{n,shot}}} = q\left(I_{\text{ph}} + I_{\text{dark}}\right) t_{\text{int}} \tag{3.14}$$

$$= q\left(Q_{\text{sig}} + I_{\text{dark}} t_{\text{int}}\right) \tag{3.15}$$

where I_{dark} is the dark current, Q_{sig} is the signal charge, and q, I_{ph} and t_{int} are defined as in equations 3.12 and 3.13.

Nyquist noise of storage capacitor

The mean-square charge uncertainty on capacitor C is

$$\overline{Q^2_{\text{n,storage}}} = kTC, \tag{3.16}$$

where k is Boltzmann's constant and T is the absolute temperature. This noise can be quite significant; however, it is suppressed by correlated double sampling, as explained above.

Source follower: Nyquist and flicker noise

The noise due to the source follower is computed in the following straightforward way[104, 105]:

1. Noise voltage sources are added to the gates of both transistors, with a power spectral density appropriate to the type of noise being modeled.

2. The voltage gains between the gates and the output are computed.

3. The power spectral densities, weighted by the square of the voltage gains, are added, yielding the total output power spectral density.

The power spectral densities of Nyquist (thermal) and flicker ($1/f$) noise for a transistor of channel width W and channel length L are respectively[104]

$$\overline{e^2_{\text{n,Nyquist}}} = \frac{8kT\,(1+\eta)}{3g_m} \tag{3.17}$$

and

$$\overline{e^2_{\text{n,flicker}}} = \left(\frac{K_F}{2C_{\text{ox}}K'}\right)\frac{1}{fWL}, \tag{3.18}$$

where f is the frequency, g_m is the transconductance of the transistor, C_{ox} is the capacitance per unit area of its gate oxide, K' is its transconductance parameter and η is the strength of the body effect relative to the transconductance g_m[104]. A typical value of the constant K_F is[104]

$$K_F = 10^{-28}\text{F}\cdot\text{A}. \tag{3.19}$$

We now proceed to step 2: the computation of the voltage gain between the gate of each transistor and the output. Figure 3.18 shows the schematic of the source follower and its small-signal AC model[11]. The current sources represent the transconductance of the transistors, while the resistors represent their drain conductance. The upper supply rail V_{dd} and the bias voltage V_{SFNBIAS} are at AC ground. Small-signal AC voltages of amplitude v_{in} and v_{bn} are assigned to the gates of the input transistor Q3 and of the bias transistor Qb. Applying Kirchhoff's current law to the output node, we obtain

$$v_{\text{out}} = \frac{g_{m3}v_{\text{in}} - g_{mb}v_{\text{bn}}}{g_{m3} + g_{ds3} + g_{dsb}}. \tag{3.20}$$

The transconductance of Q3 can be assumed to be significantly larger than the drain conductances of Q3 and Qb. Therefore, the voltage gains between the gates of the input and bias transistors and the output are respectively 1 and $-g_{mb}/g_{m3}$.

[11]This simple analysis neglects the body effect.

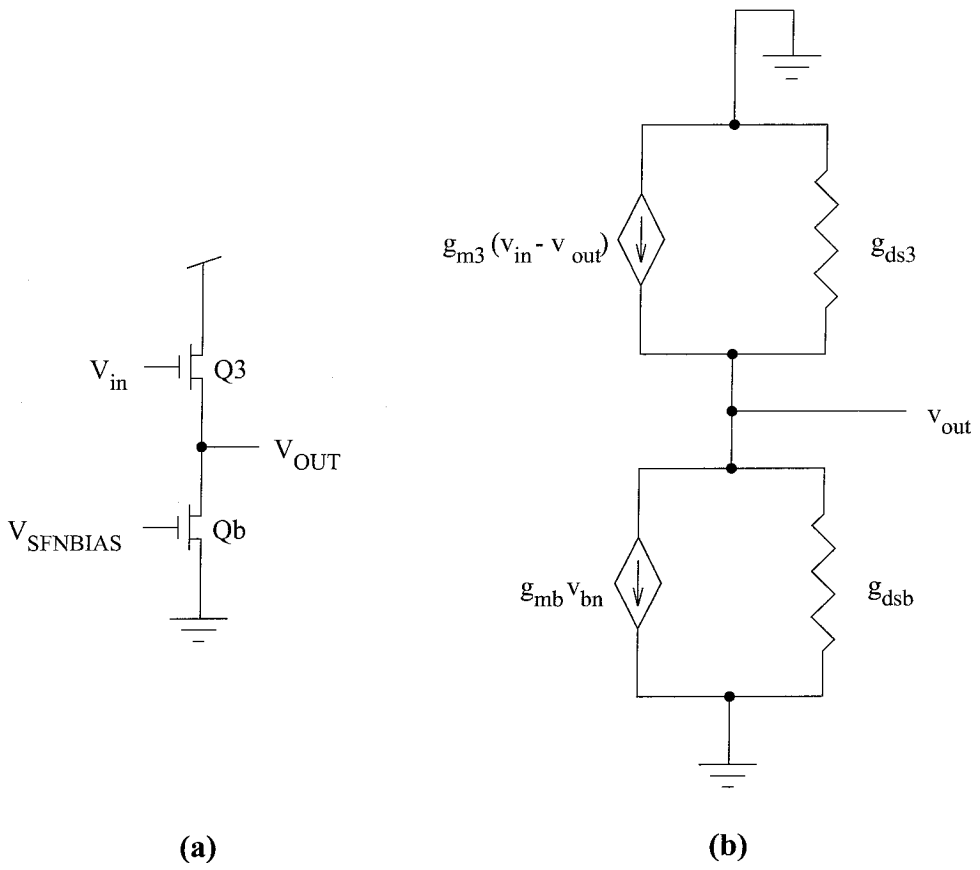


Figure 3.18: Schematic (a) and small-signal AC model (b) of the source follower.

As a third and final step, we obtain the total output power spectral density by adding the power spectral densities at the gate of each transistor, weighted by their respective squared voltage gains:

$$\overline{e^2_{\text{n,out}}} = \overline{e^2_{\text{n,Q3}}} + \left(\frac{g_{mb}}{g_{m3}}\right)^2 \overline{e^2_{\text{n,Qb}}} \tag{3.21}$$

$$= \overline{e^2_{\text{n,Q3}}}\left(1 + \left(\frac{g_{mb}}{g_{m3}}\right)^2 \frac{\overline{e^2_{\text{n,Qb}}}}{\overline{e^2_{\text{n,Q3}}}}\right), \tag{3.22}$$

where $\overline{e^2_{\text{n,Q3}}}$ and $\overline{e^2_{\text{n,Qb}}}$ are respectively the input power spectral densities of the input and bias transistors. The power spectral densities are additive because the noise sources are assumed to be independent.

We now successively apply equation 3.22 to Nyquist and flicker noise, whose input power spectral densities are given by equations 3.17 and 3.18, respectively. We will need an expression for the transconductance of a MOSFET in strong inversion and in saturation[28, 104]:

$$g_m = \sqrt{2K'\frac{W}{L}I_b}, \tag{3.23}$$

where I_b is the amplifier's bias current:

$$I_b = \frac{K'}{2}\frac{W}{L}\left(V_{\text{SFNBIAS}} - V_T\right)^2, \tag{3.24}$$

where V_T is the threshold voltage of Qb.

In the case of Nyquist (thermal) noise, we obtain

$$\overline{e^2_{\text{n,out,Nyquist}}} = \frac{8kT\left(1+\eta_3\right)}{3\sqrt{2K'_3\frac{W_3}{L_3}I_b}}\left(1 + \left(\frac{g_{mb}}{g_{m3}}\right)^2 \frac{g_{m3}}{g_{mb}}\left(\frac{1+\eta_b}{1+\eta_3}\right)\right) \tag{3.25}$$

$$= \frac{8kT\left(1+\eta_3\right)}{3\sqrt{2K'_3\frac{W_3}{L_3}I_b}}\left(1 + \sqrt{\frac{K'_bW_bL_3}{K'_3W_3L_b}}\left(\frac{1+\eta_b}{1+\eta_3}\right)\right). \tag{3.26}$$

Nyquist noise can be minimized by increasing the width to length ratio of Q3.

The output power spectral density of flicker noise is obtained in a similar fashion.

$$\overline{e^2_{\text{n,out,flicker}}} = \frac{B_3}{fW_3L_3}\left(1+\left(\frac{g_{mb}}{g_{m3}}\right)^2\frac{B_bW_3L_3}{B_3W_bL_b}\right) \tag{3.27}$$

$$= \frac{B_3}{fW_3L_3}\left(1+\frac{K'_bB_b}{K'_3B_3}\left(\frac{L_3}{L_b}\right)^2\right), \tag{3.28}$$

where, for each transistor, we define the constant B as

$$B \equiv \frac{K_F}{2C_{\text{ox}}K'}. \tag{3.29}$$

We will in general neglect the beneficial low-pass filtering effect of correlated doubled sampling; as a result, the noise level will be somewhat overestimated. We will, however, take this effect into account in the case of flicker noise, on which it has a very significant effect (recall that the flicker noise PSD goes as $1/f$).

While correlated double sampling is time-variant, White et al.[103] suggest modeling its effect on flicker noise using the following transfer function:

$$H_{\text{CDS}}(s) = \frac{1-e^{-st_{\text{int}}}}{1+s/\omega_0}, \tag{3.30}$$

where ω_0 is the bandwidth of the differential amplifier and s is the Laplace transform variable. The mean-square output noise voltage is obtained by integrating the power spectral density over the bandwidth $2/t_{\text{int}}$, with the restriction mentioned in the previous paragraph:

$$\overline{v^2_{\text{n,out}}} = \overline{v^2_{\text{n,in}}}$$

$$= \overline{e^2_{\text{n,out,Nyquist}}}\frac{2}{t_{\text{int}}} + \int_0^{2/t_{\text{int}}}\overline{e^2_{\text{n,out,flicker}}}\,|H_{\text{CDS}}(j2\pi f)|^2\,df \tag{3.31}$$

$$= \frac{16kT\,(1+\eta_3)}{3t_{\text{int}}\sqrt{2K'_3\frac{W_3}{L_3}I_b}}\left(1+\sqrt{\frac{K'_bW_bL_3}{K'_3W_3L_b}}\left(\frac{1+\eta_b}{1+\eta_3}\right)\right) + \frac{B_3}{W_3L_3}\left(1+\frac{K'_bB_b}{K'_3B_3}\left(\frac{L_3}{L_b}\right)^2\right)\int_0^{2/t_{\text{int}}}\frac{|H_{\text{CDS}}(j2\pi f)|^2}{f}\,df. \tag{3.32}$$

The mean-square output noise of the source follower is equal to its mean-square input noise because the amplifier has unity gain[12].

We now need to express the mean-square noise voltage (or noise variance) of the source follower as an equivalent mean-square charge on the storage node of capacitor C. Since $Q_{\text{storage}} = CV_{\text{in}}$,

$$\overline{Q_{\text{n,SF}}^2} = 2C^2\overline{v_{\text{n,out}}^2} \tag{3.33}$$

$$= 2C^2\frac{16kT(1+\eta_3)}{3t_{\text{int}}\sqrt{2K_3'\frac{W_3}{L_3}I_b}}\left(1+\sqrt{\frac{K_b'W_bL_3}{K_3'W_3L_b}}\left(\frac{1+\eta_b}{1+\eta_3}\right)\right) + 2C^2\frac{B_3}{W_3L_3}\left(1+\frac{K_b'B_b}{K_3'B_3}\left(\frac{L_3}{L_b}\right)^2\right)\int_0^{2/t_{\text{int}}}\frac{|H_{\text{CDS}}(j2\pi f)|^2}{f}\,df. \tag{3.34}$$

The factor of 2 in equation 3.33 is due to the statistically independent noise contributions caused by the two successive uses of the amplifier: one to read the reset level, and one to read the accumulated photogenerated charge.

Nyquist noise of the sample-and-hold circuit

Charging the capacitors of the sample-and-hold circuit causes an equivalent mean-square input noise voltage of

$$\overline{V_{\text{n,SH}}^2} = kT\left(\frac{1}{C_S}+\frac{1}{C_R}\right). \tag{3.35}$$

The equivalent mean-square charge noise on the storage node of capacitor C is

$$\overline{Q_{\text{n,SH}}^2} = kTC^2\left(\frac{1}{C_S}+\frac{1}{C_R}\right). \tag{3.36}$$

Noise due to the difference amplifier

The noise due to the difference amplifier depends on the details of its implementation. However, in general, it is possible to minimize both Nyquist and flicker noise by an appropriate design of the input transistors[104, §§6.1 and 6.2, for example]. Given area

[12]This is an approximation, which neglects the effect of the body effect in Q3 and of the drain conductances of both Q3 and Qb.

constraints and array density requirements, the difference amplifier offers considerably more device scaling latitude than the pixel circuits. Therefore, we will simply assume that the input MOSFETs of the difference amplifier can be shaped in such a way that the amplifier does not significantly increase the input-referred noise.

Total noise referred to the storage node

We represent the total noise as an equivalent mean-square charge on capacitor C:

$$\overline{Q_{\mathrm{n}}^2} = \overline{Q_{\mathrm{n,shot}}^2} + \overline{Q_{\mathrm{n,SF}}^2} + \overline{Q_{\mathrm{n,SH}}^2}. \tag{3.37}$$

The pixel kTC noise $\overline{Q_{\mathrm{n,storage}}^2}$ was not included in equation 3.37, because it is suppressed by correlated double sampling. In order to simplify the expressions, we make the following assumptions.

- Q3 and Qb possess the same structure, and thus the same physical constants (K', B, η);
- we can neglect the body effect ($\eta_3 \ll 1$, $\eta_b \ll 1$);
- the two storage capacitors of the differential sample-and-hold circuit have identical geometry and structure ($C_S = C_R$).

The total noise can then be expressed as follows:

$$\begin{aligned}\overline{Q_{\mathrm{n}}^2} &= q\left(I_{\mathrm{ph}} + I_{\mathrm{dark}}\right) t_{\mathrm{int}} + \frac{32kTC^2}{3t_{\mathrm{int}}\sqrt{2K_3'\frac{W_3}{L_3}I_b}}\left(1 + \sqrt{\frac{W_bL_3}{W_3L_b}}\right) + \\ &\quad 2C^2\frac{B_3}{W_3L_3}\left(1 + \left(\frac{L_3}{L_b}\right)^2\right)\int_0^{2/t_{\mathrm{int}}}\frac{|H_{\mathrm{CDS}}(j2\pi f)|^2}{f}\,df + \\ &\quad \frac{2kTC^2}{C_S}.\end{aligned} \tag{3.38}$$

Numerical example

We now apply the analysis presented above to estimate the sensitivity of a dynamic DHR pixel. Equations 3.38 depends on many parameters. Our assumptions are listed

Symbol	Description	Value
$W_b = W_3$	Width of transistor channels	2.1 μm
$L_b = L_3$	Length of transistor channels	2.1 μm
K'	Transconductance parameter	136 μA/V^2
C_{ox}	Gate-oxide capacitance	0.0036 F/m^2
V_T	Threshold voltage	0.63 V
K_F	Flicker noise constant	10^{-28}F $\cdot$ A
B	Lumped flicker noise constant[a]	1.02×10^{-22} m^{2}V^2
I_{dark}	Dark current	0.25 fA
C	Capacitance of storage node	55 fF
$C_S = C_R$	Sample-and-hold capacitance	3 pF
V_{SFNBIAS}	Bias voltage	1.2 V
I_b	Bias current[b]	22 μA
ω_0	Cutoff angular frequency	$2\pi/t_{\text{int}}$
η	Quantum efficiency	0.3
λ	Wavelength	670 nm

[a] $B = K_F/(2C_{\text{ox}}K')$.
[b] See equation 3.24.

Table 3.3: Parameters employed in the estimation of the dynamic pixel's sensitivity.

in Table 3.3. Almost all our process-dependent parameter assumptions are based on experimental data obtained from MOSIS characterizing advanced commercial CMOS processes. Room temperature ($T = 295$ K) is assumed. The sample-and-hold capacitance $C_S = C_R$ will be seen to play a determining role in the optical sensitivity of the circuit, which improves with increasing C_S. This is why we chose a very large value for C_S, compared to typical integrated circuit arrayed capacitances.

We now determine the minimum photogenerated signal charge, Q_{sig}, necessary to achieve a certain signal-to-noise ratio

$$\text{SNR} \equiv \frac{Q_{\text{sig}}^2}{\overline{Q_{\text{n}}^2}}. \tag{3.39}$$

To this end, the definition of a signal-independent mean-square noise charge will be

helpful:

$$\begin{aligned}\overline{Q^2_{\text{n,s-i}}} &\equiv \overline{Q^2_{\text{n}}} - qI_{\text{ph}}t_{\text{int}} \\ &= \overline{Q^2_{\text{n}}} - qQ_{\text{sig}} \\ &= qI_{\text{dark}}t_{\text{int}} + \frac{32kTC^2}{3t_{\text{int}}\sqrt{2K'_3\frac{W_3}{L_3}I_b}}\left(1+\sqrt{\frac{W_bL_3}{W_3L_b}}\right) + \\ &\quad 2C^2\frac{B_3}{W_3L_3}\left(1+\left(\frac{L_3}{L_b}\right)^2\right)\int_0^{2/t_{\text{int}}}\frac{|H_{\text{CDS}}(j2\pi f)|^2}{f}\,df + \\ &\quad \frac{2kTC^2}{C_S}.\end{aligned} \tag{3.40}$$

The minimum signal charge Q_{sig} is obtained by solving

$$\frac{Q^2_{\text{sig}}}{\text{SNR}} = \underbrace{qQ_{\text{sig}}}_{\text{Signal shot noise}} + \overline{Q^2_{\text{n,s-i}}}.$$

The solution is

$$Q_{\text{sig}} = q\,\frac{\text{SNR}}{2}\left(1+\sqrt{1+4\,\frac{\overline{Q^2_{\text{n,s-i}}}}{q^2\,\text{SNR}}}\right). \tag{3.41}$$

In general, the minimum signal power Q_{sig} is a function of the integration time, upon which depend the source follower's Nyquist and flicker noise. However, given the parameters of Table 3.3, Q_{sig} is found to be approximately independent of t_{int} over a large range of integration times,

$$25\ \mu\text{s} \le t_{\text{int}} \le 10\ \text{ms}.$$

This is because the signal-independent noise $Q^2_{\text{n,s-i}}$ is dominated by the sample-and-hold Nyquist noise,

$$\frac{2kTC^2}{C_S},$$

which is independent of the integration time. Over the range of integration times given above, the sample-and-hold Nyquist noise accounts for more than 70% of the total mean-square noise charge and more than 82% of the signal-independent mean-

square noise charge. If the minimum SNR is fixed at 9 (i.e., the signal charge must be at least three times the root-mean-square noise charge (standard deviation)), the minimum signal power varies by less than 2%. Under these conditions, the optical sensitivity depends only on the capacitances of the pixel and of the sample-and-hold circuit:

$$Q_{\text{sig}} \approx q\frac{\text{SNR}}{2}\left(1+\sqrt{1+4\,\frac{2kTC^2}{q^2C_S\,\text{SNR}}}\right), \tag{3.42}$$

$$Pt_{\text{int}} \approx \frac{hc}{\eta\lambda}\frac{\text{SNR}}{2}\left(1+\sqrt{1+4\,\frac{2kTC^2}{q^2C_S\,\text{SNR}}}\right), \tag{3.43}$$

where Pt_{int} is the minimum optical exposure needed to achieve the desired signal-to-noise ratio. With the parameters listed in Table 3.3, we obtain, for SNR = 9,

$$Q_{\text{sig}} \approx 9.3\times10^{-18}\ \text{C} = 58q \tag{3.44}$$

$$Pt_{\text{int}} \approx 58\ \text{aJ}. \tag{3.45}$$

A more accurate solution including all the noise components in equation 3.40 (instead of only the sample-and-hold Nyquist noise) gives $Q_{\text{sig}} = 63q$ and $Pt_{\text{int}} = 63$ aJ at $t_{\text{int}} = 1$ ms. The design of a compact, low-noise, low-offset difference amplifier suitable for such low-exposure detection may be rather challenging. The signal voltage swing is only 0.2 mV at $Q_{\text{sig}} = 63q$. In practice, a larger signal charge, resulting in a larger SNR, is likely to be required.

Consider the case where the power incident on each pixel is $P = 360$ fW and the integration time[13] is $t_{\text{int}} = 1$ ms. This is the signal power than would be incident on each photodetector in an optical memory system with an effective $M/\# = 4$ multiplexing 4096 holograms, each containing 1.5 K × 1.5 K = 1536 × 1536 pixels[14]

[13]We limit the integration to 1 ms to ensure that it is comparable to the settling time of a nematic liquid crystal beam steerer, and therefore that it does not unduly increase the access time of the memory.

[14]1 K = 2^10 = 1024.

Description	mean-square noise charge	RMS noise
Signal shot noise	9.3×10^{-36} C^2	$19q$
Dark current shot noise	4.0×10^{-38} C^2	$1.2q$
Source-follower Nyquist noise	3.4×10^{-39} C^2	$0.4q$
Source-follower flicker noise	1.3×10^{-36} C^2	$7.2q$
Sample-and-hold Nyquist noise	8.2×10^{-36} C^2	$17.9q$
Total noise	1.9×10^{-35} C^2	$27.1q$
Pixel Nyquist (kTC) noise[a]	2.2×10^{-34} C^2	$93q$

[a] Correlated double sampling suppresses this noise, which is thus not included in the total noise.

Table 3.4: Noise breakdown at $P = 360$ fW and $t_{\text{int}} = 1$ ms.

(these numbers will seem more meaningful in Chapter 5). The signal charge is then

$$Q_{\text{sig}} = 368q.$$

The root-mean-square total noise, obtained from equation 3.38, is

$$\sqrt{\overline{Q_{\text{n}}^2}} = 27.1q,$$

and the optically induced voltage swing at the storage node is 1.1 mV. Table 3.4 gives the individual noise contributions. The sensitivity is limited by signal shot noise and by the sample-and-hold Nyquist noise. We can appreciate the importance of correlated double sampling in suppressing the pixel kTC noise, which would otherwise dominate the noise performance of the circuit.

Chapter 4 Neural arrays

In §§1.1.5 and 1.2, we argued that the nonlinear neuron responses in optoelectronic neuromorphic systems are best implemented using optoelectronic components, and we saw that the most common such responses are threshold and "bump" functions. Optoelectronic integrated circuits are particularly well suited to the implementation of neural responses because they combine the flexibility of conventional VLSI circuits with optical inputs and outputs.

This Chapter describes two neural arrays in which each neuron either performs thresholding or provides a bump response. It is brief for two reasons. First, the material presented in it is not used in subsequent Chapters. Second, the devices have not yet been fully tested and characterized. Readers who are primarily interested in holographic data storage may want to skip to Chapter 5.

The first device, described in §4.1, employs GaAs multiple-quantum-well (MQW) modulators bonded to a silicon die, using the Hybrid-SEED process developed by Bell Laboratories[33]. §4.2 discusses a GaAs OEIC using MESFET detectors and light-emitting diodes (LEDs) grown by means of molecular-beam epitaxy after the fabrication of the rest of the integrated circuit[30].

The functional diagram of a pixel (neuron) of either array is shown in Figure 4.1. The pixel receives a differential optical input signal, computes the desired bump or threshold function, and provides an optical output. Both arrays were designed to also operate with single-ended optical input signals, if desired. In the case of the multiple-quantum-well device, optical detection is effected by GaAs multiple-quantum-well p-i-n diodes, the response is computed by a silicon VLSI circuit, and a differential optical output is provided by a pair of multiple-quantum-well modulators. The GaAs device detects input signals using MESFET optical detectors (OPFETs)[106, Chapter 4] and LED outputs.

Access to both technologies was provided by the Consortium for Optical and Op-

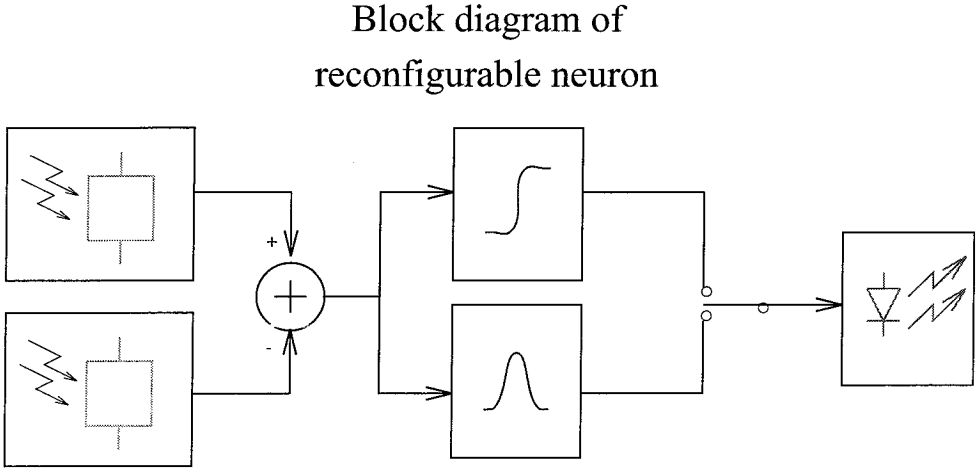


Figure 4.1: Functional diagram of a threshold/bump pixel.

toelectronic Technologies for Computing (CO-OP) under the direction of Dr. Ravi Athale and Dr. Kannan Raj, to whom we express our gratitude. We are also thankful to the technology providers: Bell Laboratories in the case the the Hybrid-SEED process, and the Massachusetts Institute of Technology (Dr. Clifton Fonstad and Joe Ahadian) in the case of the OPTOCHIP LED regrowth project.

4.1 Hybrid-SEED neural array

4.1.1 Structure and properties of Hybrid-SEED modulators

This OEIC employs the Hybrid-SEED process[33] developed by Lucent Technologies' Bell Laboratories (while the Laboratories still belonged to AT&T). The process allows GaAs MQW-SEED modulators to be bonded with high yields to conventionally processed silicon CMOS VLSI die. This approach offers many advantages compared to other modulator technologies and previous monolithic SEED implementations. Hybrid-SEED circuits offer fast circuit-limited[1] light modulation, the flexibility and yields of conventional CMOS VLSI circuits, and 3-D integration of modulators and information processing circuits, because the GaAs modulators can be bonded on

[1] As opposed to modulator-limited

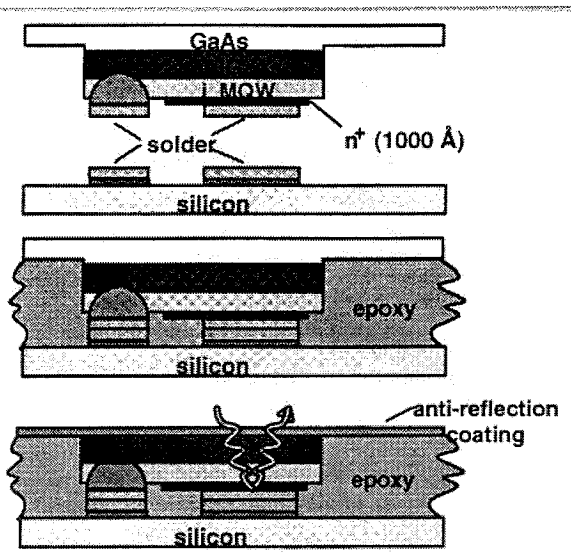


Figure 4.2: Structure of a Hybrid-SEED modulator.

top of the silicon circuits that control them. The use of reflective modulators offers several advantages:

- contrast is improved because of the two passes of readout beams through the electronically controlled absorptive modulator;
- an opaque substrate can be employed;
- the OEIC is easily packaged, mounted, and fitted with a heat sink.

High yields (fraction of non-working devices less than 0.001) were already routinely obtained at the time of the CO-OP SEED'95 workshop in July 1995.

Figure 4.2 shows a cross-section schematic of a Hybrid-SEED modulator. A GaAs die, comprising several multiple-quantum-well diodes grown by molecular-beam epitaxy, is flip-chip bonded to a silicon die containing all the information processing circuitry. The attachment and electrical contacts are provided by thermal compression bonds. The cavity between the die is filled with epoxy. The GaAs substrate is then removed. Finally, an anti-reflection coating is deposited on the surface of the OEIC. Only the topmost metal level of the silicon IC is used as a bonding pad;

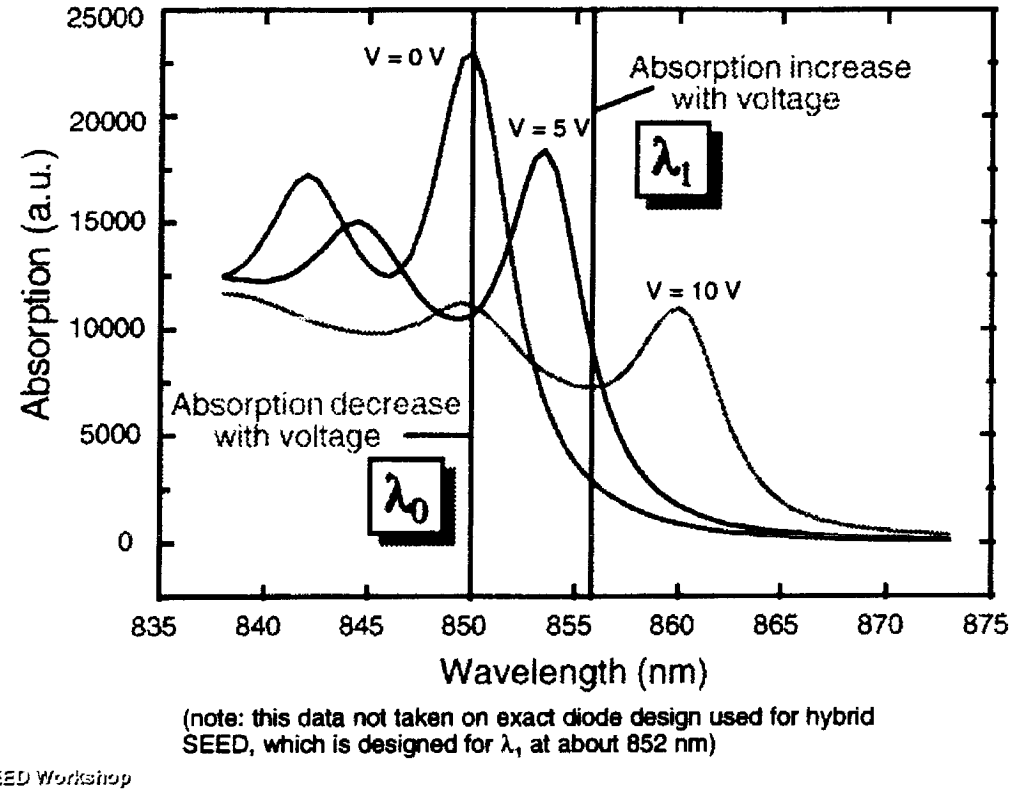


Figure 4.3: Electrooptics response of a MQW modulator (experimental data obtained from Bell Laboratories).

interconnections and circuit elements can reside under the pad.

The electrooptic response of a MQW modulator is shown in Figure 4.3. When an electric field is applied across a multiple-quantum well structure, the minimum photon energy at which significant absorption occurs is decreased. Our modulators are designed to operate in the so-called λ_1 mode, i.e., at a wavelength slightly longer than that of the exciton peak. A modulator absorbs only a small fraction of the incident light when no voltage is applied across it. The absorption increases with increasing voltage, until the exciton peak is shifted to the operating wavelength.

The λ_1 mode lends itself to negative feedback and enables an elegant self-linearized modulation method[107], illustrated in Figure 4.4a. The current I_c controls the light

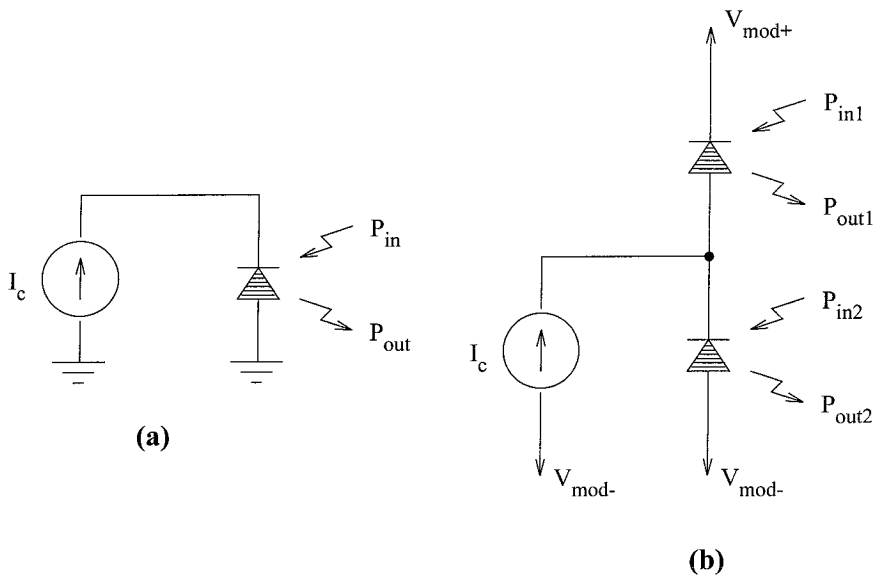


Figure 4.4: Self-linearized single-ended (a) and differential (b) modulators.

power subtracted from the input beam by the MQW p-i-n diode. Let us assume that

$$\frac{P_{\text{in}} q\lambda}{hc} > I_c,$$

where q is the charge of an electron, λ is the wavelength of the readout beam, h is Planck's constant and c is the speed of light in vacuum; in other words, the control current I_c is less than the photocurrent at 100% absorption. We also assume that the MQW diode gives one electron per absorbed photon. The condition

$$\frac{(P_{\text{in}} - P_{\text{out}})\, q\lambda}{hc} = I_c, \tag{4.1}$$

is stable: the photocurrent is compatible with the current imposed by the current source. Suppose that equation 4.1 is initially satisfied, and that I_c is then slightly increased. Initially, the MQW diode's reverse bias increases, and, in turn, its absorption increases (see Figure 4.3). The photocurrent produced by the diode increases until equation 4.1 is once again satisfied, at the new value[2] of I_c. Therefore, *the absorbed*

[2]Any absorption overshoot would be corrected, because the excessive photocurrent would decrease the MQW diode's reverse bias, and consequently decrease its absorption.

power is proportional to the control current I_c.

In practice, because of their relatively low contrast ratios (2:1 to 5:1), MQW diodes are often used in pairs, with differential optical signals. The self-linearized mode described above also applies to this arrangement (see Figure 4.4b). In this case, I_c is proportional to the difference in absorbed power between the two modulators. Note that, in order to cover the full differential range of the pair of modulators, a symmetric range of positive and negative currents I_c must be available.

4.1.2 Neuron circuit

Our neuron circuit, which is designed to operate around $\lambda = 852$ nm, uses self-linearized modulators for its optical outputs. A pair of optical receivers transforms the input photocurrents obtained from two p-i-n photodiodes into voltages that drive threshold and bump circuits.

Optical receiver

The optical receiver consists of a transimpedance amplifier with a feedback path containing a diode-connected n-channel MOSFET and a p-channel MOSFET with its gate grounded (see Figure 4.5)[108, 109]. The transistors were sized to ensure a relatively smooth input-output characteristic with an output voltage range significantly greater than 0.4 V, which is approximately the full width at half maximum of the bump function (as we will see in Figure 4.8). The input-output characteristic of the receiver circuit is shown in Figure 4.6. It was measured on a device without MQW diodes. An externally injected current was substituted for the MQW diode's photocurrent. Based on this characteristic, we estimate that the input gain of the circuit will be -32 mV/μW.

Each neuron contains two optical receivers, one for each component of its differential optical input. Single-ended operation is also supported. In this case, SE is asserted, and an externally supplied voltage, V_{SEIN}, is substituted for the output of one of the receivers.

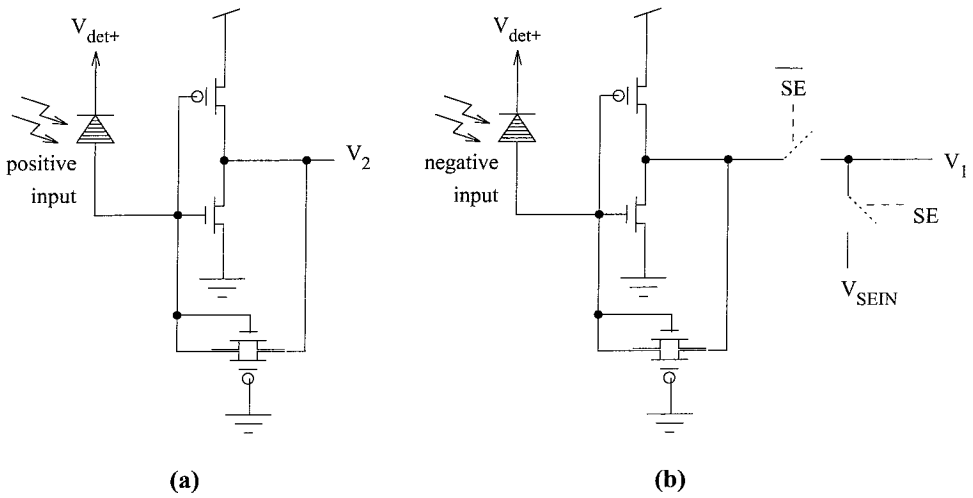


Figure 4.5: Schematic of optical receivers.

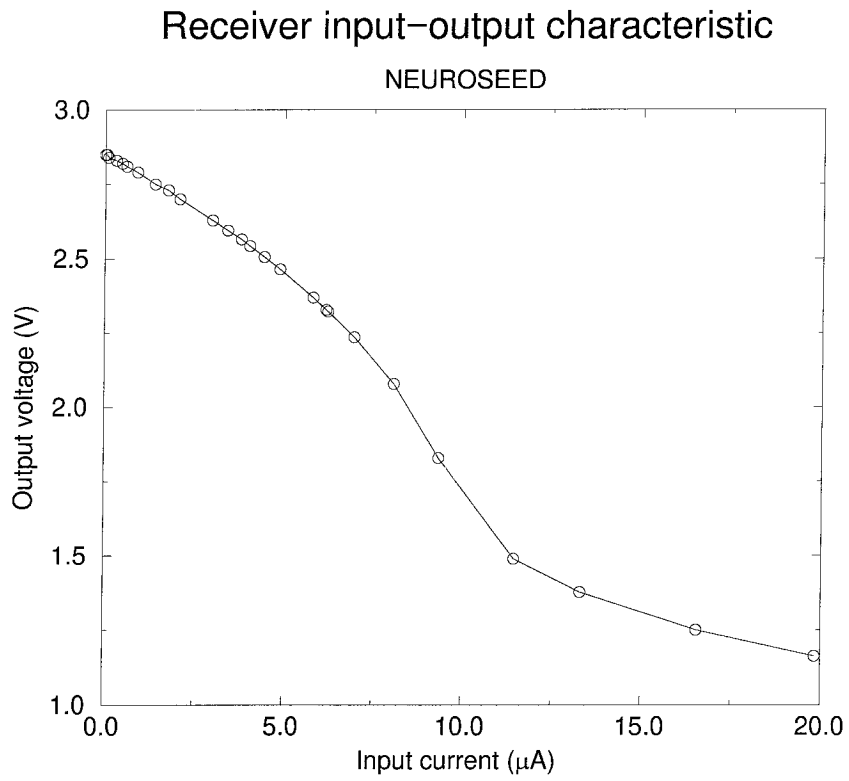


Figure 4.6: Input-output characteristic of optical receiver.

Computation circuit

The computation circuit, whose schematic is shown in Figure 4.7, is comprised of a transconductance amplifier[98] (top of Figure 4.7), a bump-anti-bump circuit[110] (bottom of Figure 4.7) and an analog multiplexer that allows either of these two circuits to drive, by means of its output current, a pair of MQW modulators in a differential configuration identical to that shown in Figure 4.4b. Two transistors were added to the central branch of the regular bump-anti-bump circuit[110] to improve its symmetry in the strong inversion regime (i.e., above threshold). We recall from §4.1.1 that the node driving the pair of MQW modulators must be able to source or sink current. The output of the regular transconductance amplifier possesses this property[98]. However, the two outputs of the regular bump-anti-bump circuit[110] can only sink current. The current mirrors above and to the right of the bump circuit scale the bump and anti-bump responses appropriately, and subtract the anti-bump response from the bump response. The response of the circuit was again measured on an OEIC without modulators, at[3] $V_{\mathrm{BBIAS}} = 1.0$ V. It is shown in Figure 4.8. The contrast of the modulators can be optimized by adjusting the bias voltages of the transconductance amplifier and of the bump-anti-bump circuit.

[3]The threshold voltage was approximately 0.63 V.

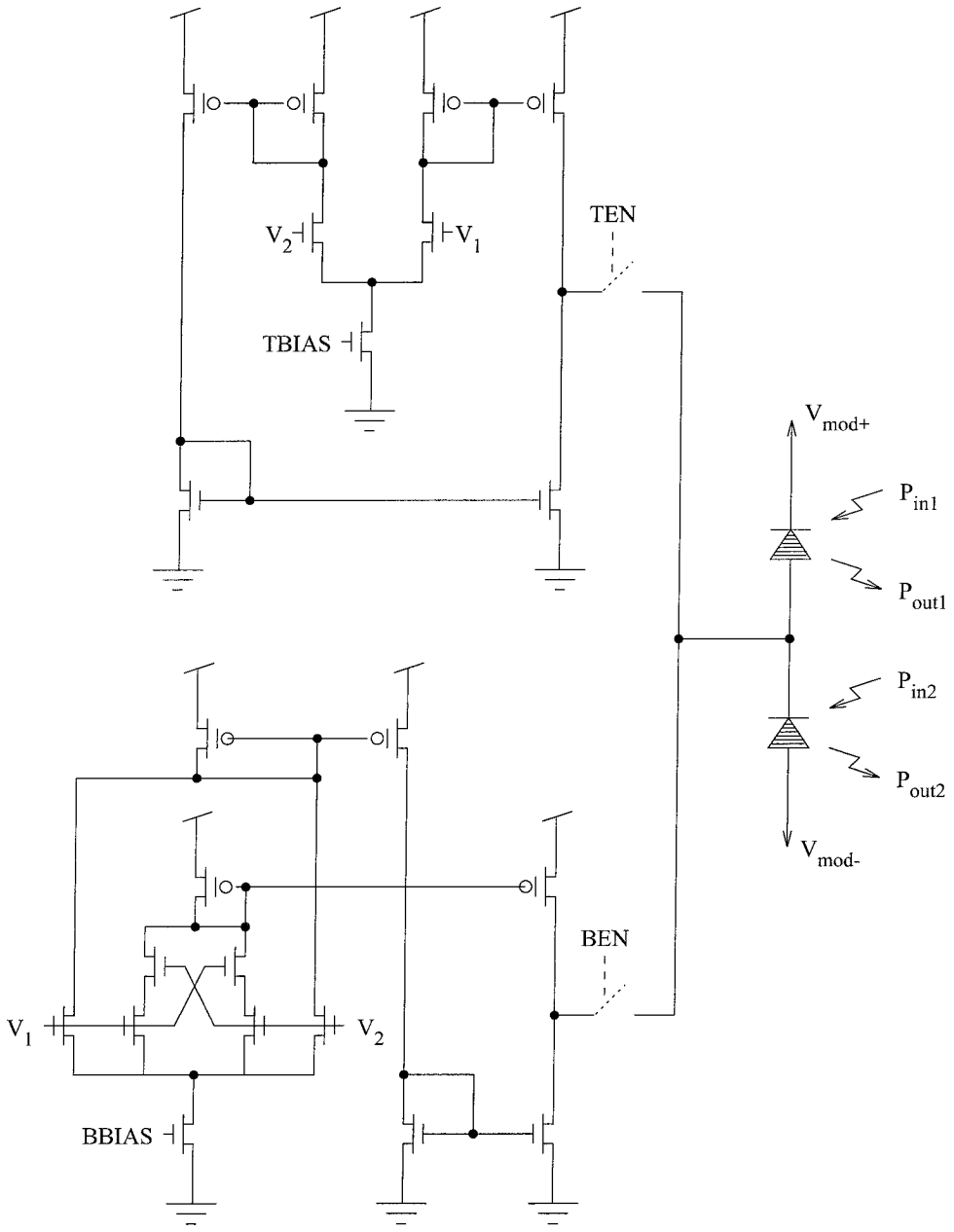


Figure 4.7: Schematic of the computation circuit.

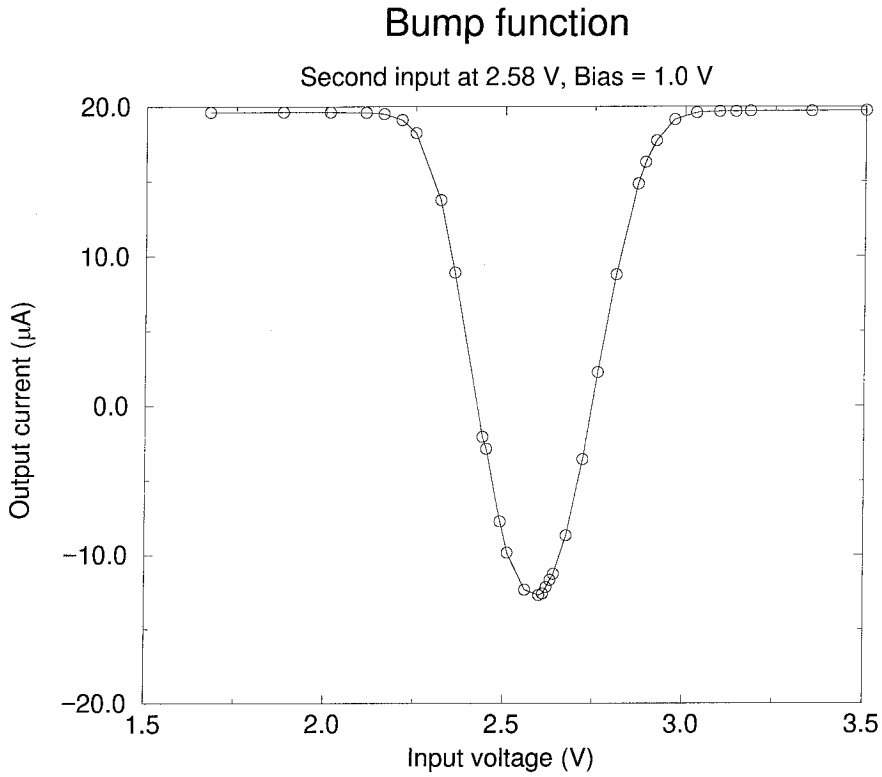


Figure 4.8: Bump response.

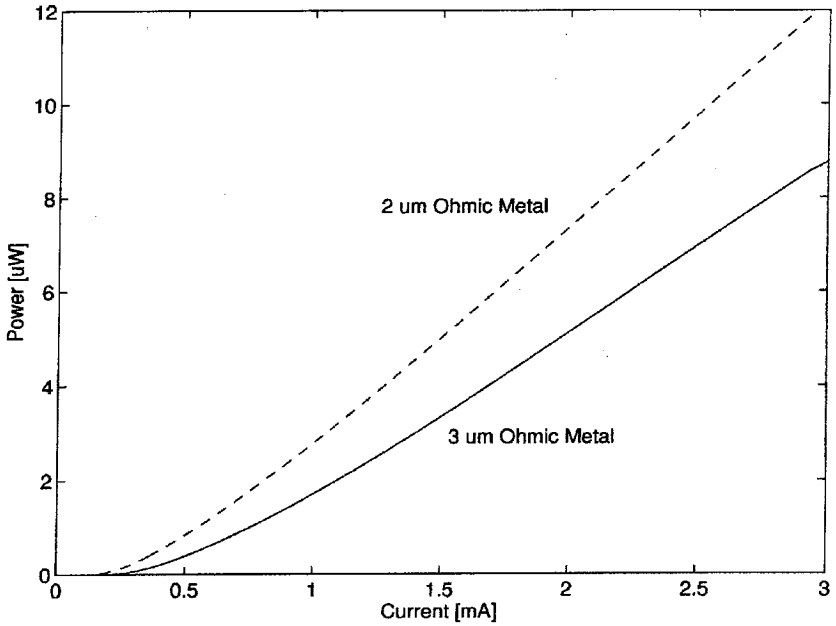


Figure 4.9: Light vs. current (LI) characteristic of an LED employing an earlier version of the same regrowth process as that used on our GaAs neural array (Experimental data obtained from MIT).

4.2 GaAs neural array with LED outputs

4.2.1 Process description

The second OEIC described in this Chapter is based on the OPTOCHIP project, directed by Dr. Clifton Fonstad at the Massachusetts Institute of Technology and sponsored by CO-OP. Light-emitting diodes are grown by molecular-beam epitaxy[30, 31] on a GaAs die fabricated using Vitesse Semiconductor's H-GaAs-3 process, which offers enhancement and depletion MESFETs.

Figure 4.9[4] shows the LI (light vs. current) curve of LED prototypes developed at MIT. The LEDs grown on our device are expected to possess equal or better performance. The emission spectrum is centered at 873 nm, which corresponds to the room-temperature bandgap of GaAs.

[4]The ohmic metal dimension mentioned in the graph is the width of the ohmic metal used to make an annular contact on top of the LED.

4.2.2 Circuit

Figure 4.10 shows the schematic of the circuit contained in each pixel (neuron). The differential optical input is sensed by a pair of optical FETs[106, 32] (OPFETs) connected in series. The two devices form a current comparator, which is not entirely unlike that consisting of Q3 and Q4 in Figure 3.5. The output of this comparator goes directly into a bump circuit and a threshold circuit, the latter being a simple inverter with an enhancement MESFET input and a depletion MESFET pull-up. The bump circuit we employed was originally proposed by our colleague Jiafu Luo[106, Figure 6.11]. The rest of the circuit consists of buffering circuits and direct-coupled FET logic (DCFL) that selects the desired function (threshold if $F = 0$ ($V_F < 0.1$ V) or bump if $F = 1$ ($V_F \approx 0.6$ V). The LED is driven by a wide enhancement MESFET that can sink up to 2.75 mA.

4.2.3 Simulation results

As of this writing, the LEDs are being fabricated at MIT. We now present simulation results obtained using a version of Meta Software's HSpice circuit simulator employing custom MESFET models developed by Vitesse Semiconductor. In these simulations, the analog front-end power supply V_{FEP} was set to 0.8 V, and the DCFL power supply V_{dd} and the drain voltage of J9 were set to 2.0 V. Figures 4.11 and 4.12 show the threshold ($F = 0$) and bump ($F = 1$) responses of the circuit, respectively. The output is the drain current of J9. Optical inputs are simulated by biasing the gates of J1 and J2. In both cases, the gate voltage of J2 is scanned between 0 and 400 mV, at several different values of the gate voltage of J1, starting at 500 mV, in 10 mV increments.

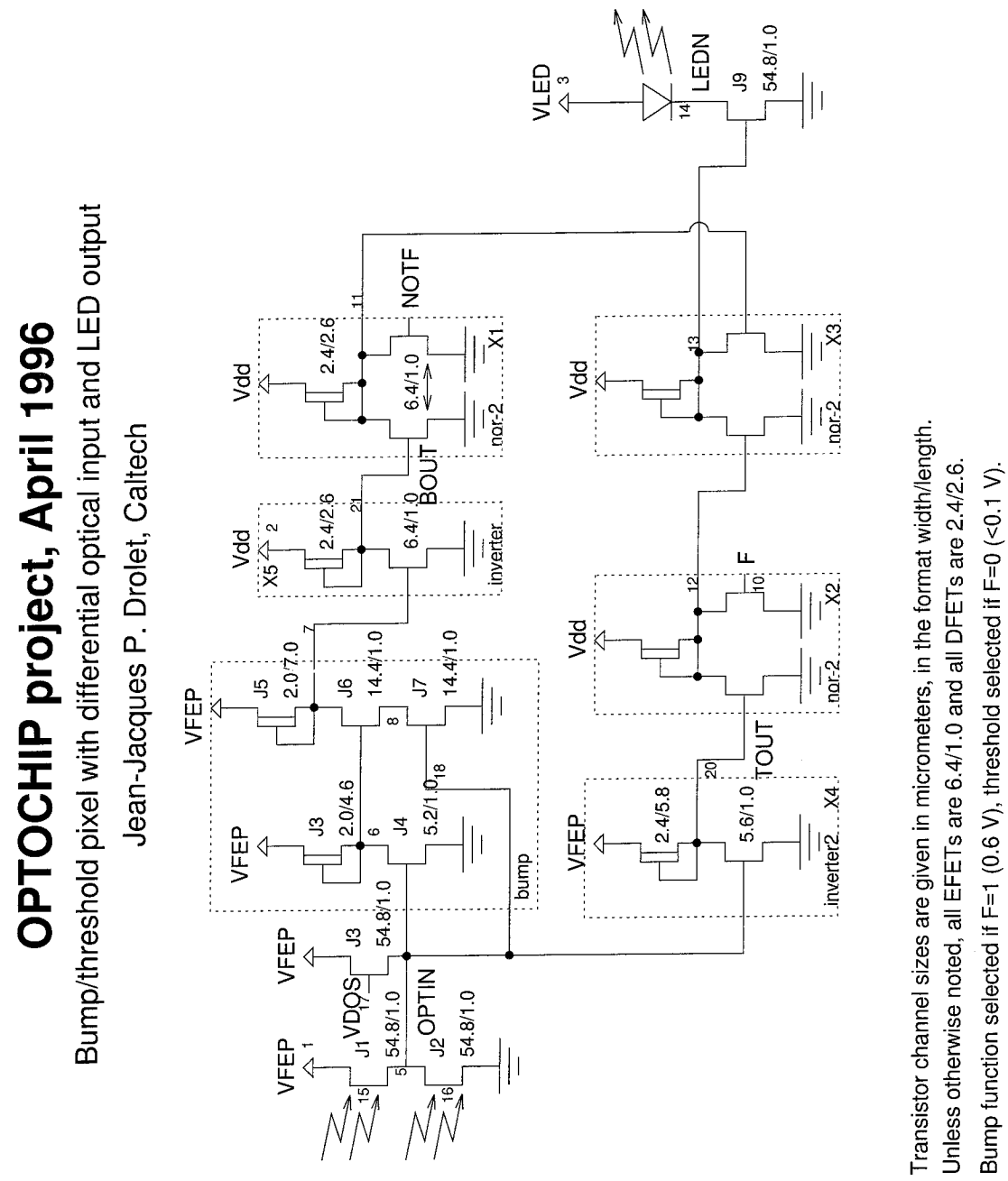


Figure 4.10: Schematic of the GaAs pixel circuit.

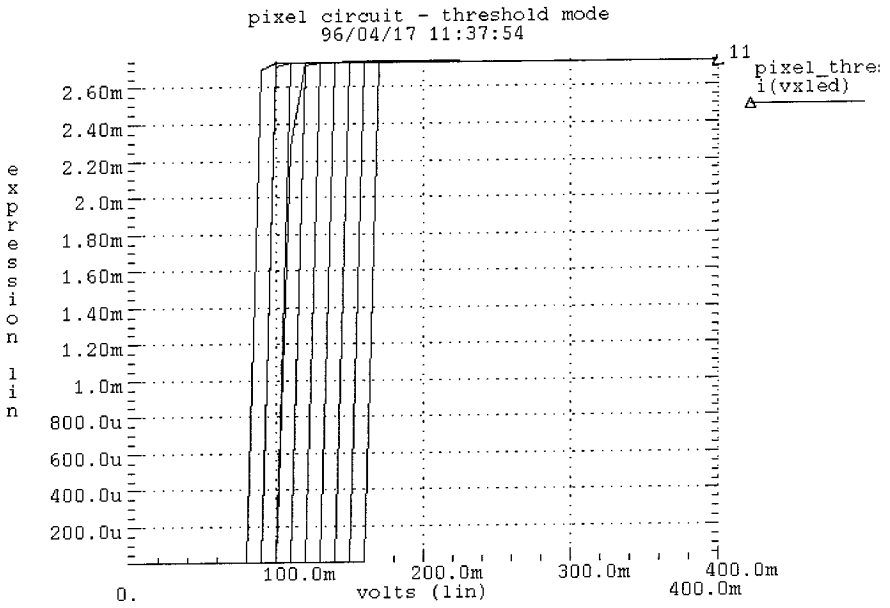


Figure 4.11: Simulation results in threshold mode.

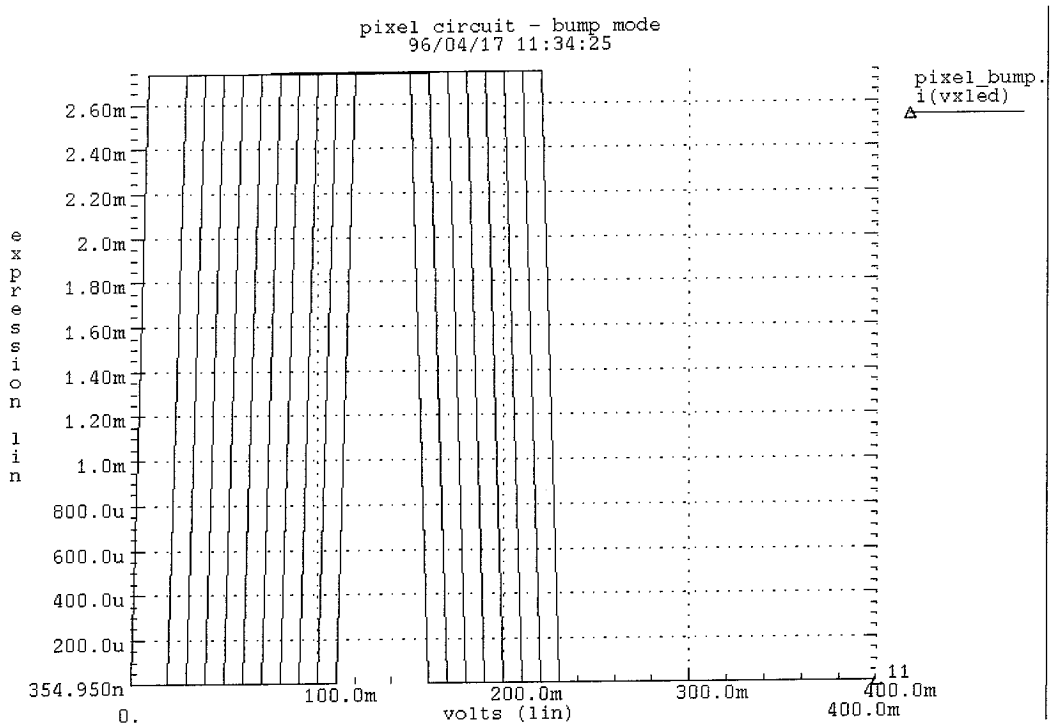


Figure 4.12: Simulation results in bump mode.

Chapter 5 Integrated holographic module architectures

5.1 Integrated modular holographic memory

We argued in Section 1.3 that compact, modular and relatively inexpensive holographic systems were enabled by conjugate readout and by an OEIC called *Dynamic Hologram Refresher*, or DHR for short. We refer to systems assembled from compactly integrated holographic storage modules such as that shown in Figure 1.7 as *integrated modular holographic memory.* In this chapter, we specify and attempt to optimize the architectural details of integrated storage modules such as the one illustrated in Figure 1.7, and we formulate storage density and cost projections for integrated modular holographic memory.

We first present an interesting way of thinking of integrated holographic storage modules in §5.1.1. The DHR is thought of as a dynamic RAM (DRAM) with optical input/output functions built into each pixel, and the optical components of the module are presented as a means to multiply the capacity of the special DRAM chip by a large integer, namely, the number of holograms. Then, in §5.1.2, we discuss the modularity, scalability, flexibility and reconfigurability of integrated modular holographic memory. In particular, we argue that integrated modular holographic memory systems can be upgraded in the field much like DRAM single-in-line memory modules (SIMM), by adding modules as needed by the application. We also explain that a range of performance requirements can be addressed using essentially the same module architecture and hardware. §5.2 analyzes and optimizes eight variations of the basic angularly multiplexed module illustrated in Figure 1.7. Optimum storage cost and density are estimated, and compared to the cost and density of DRAM and magnetic storage in §5.3. In §5.4, we describe an elegant system for correlation (pattern recognition)

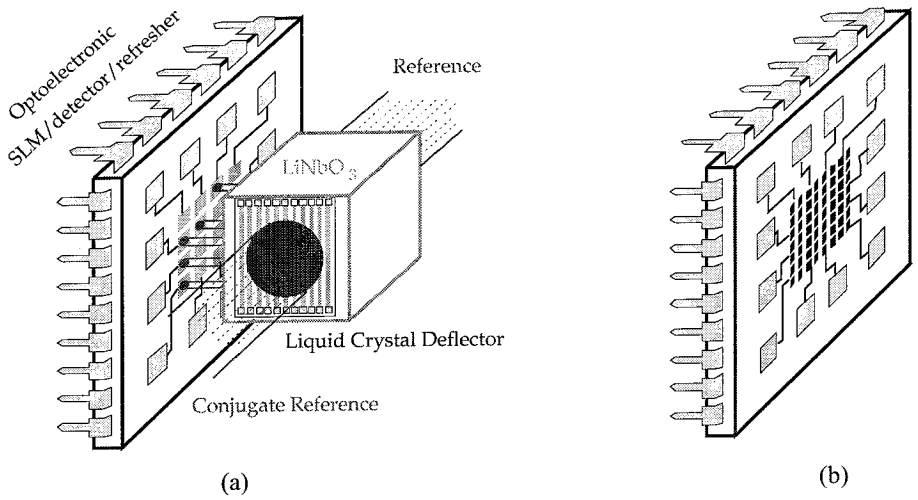


Figure 5.1: Comparison between a holographic module employing a DHR OEIC (a) and a normal RAM IC (b).

and associative memory based on conjugate readout and the DHR OEIC, and thereby justify the claim made in §1.3 that our integrative approach is applicable to systems other than simple holographic memory. Finally, §5.5 illustrates the generality of the integrative approach, by describing integrated holographic storage modules based on multiplexing techniques other than angular multiplexing.

5.1.1 Holographic storage as DRAM capacity enhancer

The Dynamic Hologram Refresher can be viewed as a memory integrated circuit in which each memory cell is endowed with optical input and output capability. This viewpoint applies equally well to the DHR-4 device (which uses a static memory circuit) as to the more recent pixel design described in §3.3, which is more akin to a dynamic RAM. Figure 5.1 schematically illustrates this comparison. It shows an integrated holographic memory module using a DHR OEIC (Figure 5.1a), compared to a regular RAM IC (Figure 5.1b). Both the DHR OEIC and the RAM IC consist of a rectangular array of pixels surrounded by control and signal processing circuitry. Let N^2 be the number of pixels in the RAM IC.

The optical input and output functions can cause a DHR pixel to incur a significant

area overhead compared to a regular memory cell. For example, if the DHR device employs LCOS modulators, the simplest pixel architecture requires that spatially disjoint pixel areas be devoted to optical input (e.g., a photodiode) and optical output (a modulator electrode) functions. Depending on the properties of the VLSI process and on the extent of the post-processing performed to fabricate the modulators, some or all of the pixel electronics can be located under the modulator. We denote the ratio of the DHR pixel area to the standard RAM cell area as R:

$$R \equiv \frac{A_{\text{DHR pixel}}}{A_{\text{RAM cell}}}.$$

The increased area of the storage cell is obviously detrimental to the holographic module. However, it is offset by the large storage capacity of the holographic medium. Let M be the number of holograms stored in the recording medium. While the standard RAM IC stores a single 2-D data page, the holographic module stores M 2-D (smaller) pages in its holographic medium. In the latter case, the integrated circuit is more a page cache than the main storage medium.

If the page size of the standard RAM IC is N^2, the number of pixels that can be fabricated in the same area on a DHR die is N^2/R. The total capacity of the RAM IC is N^2, while that of the holographic module is the DHR page size, N^2/R, times the number of holograms, M. Hence, we can say that the addition of a holographic module to a memory IC *multiplies its capacity by* M/R.

Of course, a number of other parameters are significant besides the storage capacity associated with a single IC. The following are among the most important:

- storage cost;
- storage density;
- readout rate;
- recording rate;
- access time.

Parameter	RAM IC	Holographic module
Page size	N^2	N^2/R
Number of pages	1	M
Capacity	N^2	$N^2\frac{M}{R}$
Storage cost per bit	$\frac{C_{\text{Si}}}{N^2}$	$\left(C_{b,\text{DRAM}} + \frac{C_{\text{optics}}}{N^2}\right)\frac{R}{M}$

Table 5.1: Parameters of holographic module and RAM IC.

We will consider the storage cost, C_b, in \$/bit, as an example. As a first approximation, we can assume that the areal costs of the RAM and DHR ICs are approximately equal. The storage cost of the RAM IC is

$$C_{b,\text{DRAM}} = \frac{C_{\text{Si}}}{N^2}, \tag{5.1}$$

where C_{Si} is the cost of the silicon. We argued above that the capacity of a DHR IC using the same cell array area is N^2/R bits. If we assume that the area devoted to control functions is either negligible or comparable in both integrated circuits, then the area of a DHR device caching N^2/R bits is identical to that of the RAM IC considered above. The storage cost of a holographic module employing this DHR OEIC is

$$C_{b,\text{holo}} = \frac{C_{\text{Si}} + C_{\text{optics}}}{\frac{N^2}{R}M} \tag{5.2}$$

$$= \left(C_{b,\text{DRAM}} + \frac{C_{\text{optics}}}{N^2}\right)\frac{R}{M}, \tag{5.3}$$

where C_{optics} is the cost of all of the optical and optoelectronic components (including the storage medium, the laser, the optical modulators and beam steerers, and any other support optics) of the module. Figure 5.1 summarizes these simple results.

If integrated modular holographic memory is to be cost-competitive with RAM, we must at least have $C_{b,\text{holo}} \leq C_{b,\text{DRAM}}$. The ratio R/M, which multiplies the storage cost of the RAM IC in equation 5.3, is most likely smaller than unity. The remaining

question is to what extent the added cost of the optical components,

$$\frac{C_{\text{optics}}}{N^2}\frac{R}{M},$$

offsets the scaling of the RAM storage cost by R/M. In other words, does the capacity enhancement afforded by the optical and optoelectronic portion of the module warrant its cost overhead? For example, in the case where $C_{b,\text{DRAM}} = \$7 \times 10^{-8}/\text{bit} = \$0.587/\text{MB}$, $N = \sqrt{256 \times 2^{20}}$ [111, projections for 1998], $R = 30$ and $M = 512$, the break-even point is at $C_{\text{optics}} = \$302$. *A priori,* it would seem that this cost is not overreaching, and that integrated modular holographic memory stands a good chance of being cost-competitive with RAM ICs.

However, we must temper this initial optimism with several notes of caution. The simple cost analysis presented above, like its density and performance counterparts, is a good introduction to the analysis required to estimate the storage cost of integrated modular holographic memory and to determine its competitiveness. Nevertheless, several important facts are not readily apparent in it. For example, C_{optics} depends rather strongly on N, M, and on other factors such as the available laser power and the exact details of the module architecture. In modules employing low-cost laser diodes, M is usually limited by the available power, by N, by the integration time and by the detection noise floor. Likewise, when we substituted (5.1) into (5.2) to obtain (5.3), we implicitly assumed that the areal cost of the DHR is the same as that of a standard RAM IC. This assumption may not be entirely accurate, however, because a dense DHR design is likely to require more fabrication masks, and because its production volume will be lower, at least initially. Moreover, the electronic page buffer and the control signal processing circuitry that support the pixel array in the DHR require more area than the peripheral circuitry in a RAM IC.

In general, the various parameters of the holographic memory module (i.e., number of pixels per page, pixel pitch, number of holograms, available laser power, minimum detectable power, integration time...) are deeply intertwined with each other, with technological limitations (e.g., sensitivity of recording material, operating wavelength,

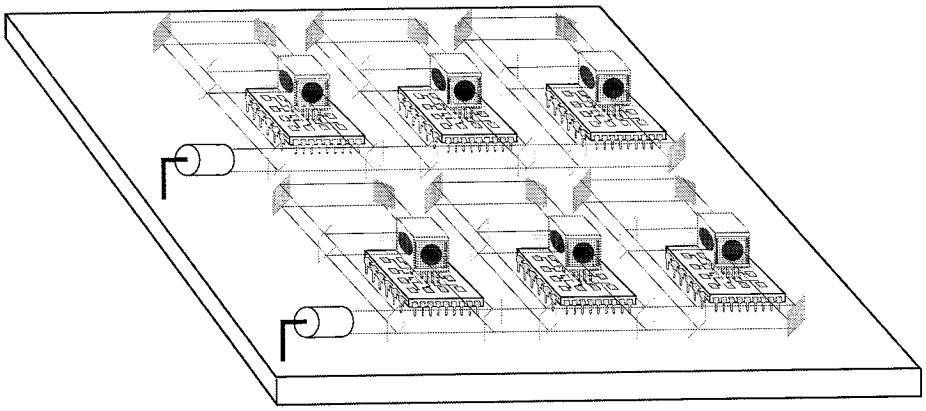


Figure 5.2: A holographic storage board, with optical components overlying the integrated circuits and printed circuit board.

maximum die size...), with cost scaling (e.g., dependence of the cost of the recording medium on the data beam size and on the reference angular range) and with the performance metrics listed above: storage cost, storage density readout rate, recording rate and access time. A careful analysis is required to properly account for all these factors. Such an analysis will be presented in §5.2. Module architectures will be optimized for storage density and cost, and will be compared to both semiconductor and magnetic storage technologies.

5.1.2 Modularity, scalability, flexibility and reconfigurability

Compact holographic modules similar to that illustrated in Figure 1.7 can be assembled in flexible, reconfigurable storage boards (see Figure 5.2). The assembly contains an electronic circuit board that interconnects the modules' OEICs, as well as an optional high-speed parallel-to-serial fiber channel interface, and an optical overlying layer containing the storage crystals, beam steerers, lasers and interconnection optics.

The modularity of this system makes it easy to tailor to specific storage requirements, and to upgrade in the field. Moreover, a wide range of performance requirements can be met by connecting the same basic storage modules to different boards.

For example, a holographic storage board employed in a cost-sensitive application might access only one or a handful of storage modules at a time and provide 16- to 64-bit wide parallel electronic outputs. A board targeted at a high-performance application (e.g., parallel supercomputers), however, could contain a 16×16 array of semi-concurrently accessible modules and provide its output on a parallel fiber channel.

5.2 Cost and density analyses

In what follows, we analyze the cost and the density of holographic storage modules employing the principles set forth in §§1.3 and 5.1. In §5.2.1, we propose four different practical implementations of the basic angularly multiplexed module architecture schematically shown in Figure 1.7. The cost and density of each module is determined by the properties of the recording material, the available optical power, the angular range of the beam steering unit, the size of the data pages (number of pixels, pixel pitch, fill factor), and the responsivity and noise properties of the photodetectors and the associated circuitry that detect optical data pages, in addition to architectural considerations.

5.2.1 Practical implementations of a holographic data storage module

We will analyze and compare the storage density and cost of the four module architectures illustrated in Figure 5.3. Each module architecture is designed to provide a signal beam and two reference beams (main reference and conjugate reference), and requires a single collimated input beam. The first two modules, illustrated in Figures 5.3a and 5.3b, employ a single beam steering unit. The conjugate readout beam, used to read out data pages, is obtained from the reflection of the main reference beam by a metallized face at the bottom of the storage crystal. This eliminates the cost and volume of two beam splitters and one beam steering unit, at the expense of a

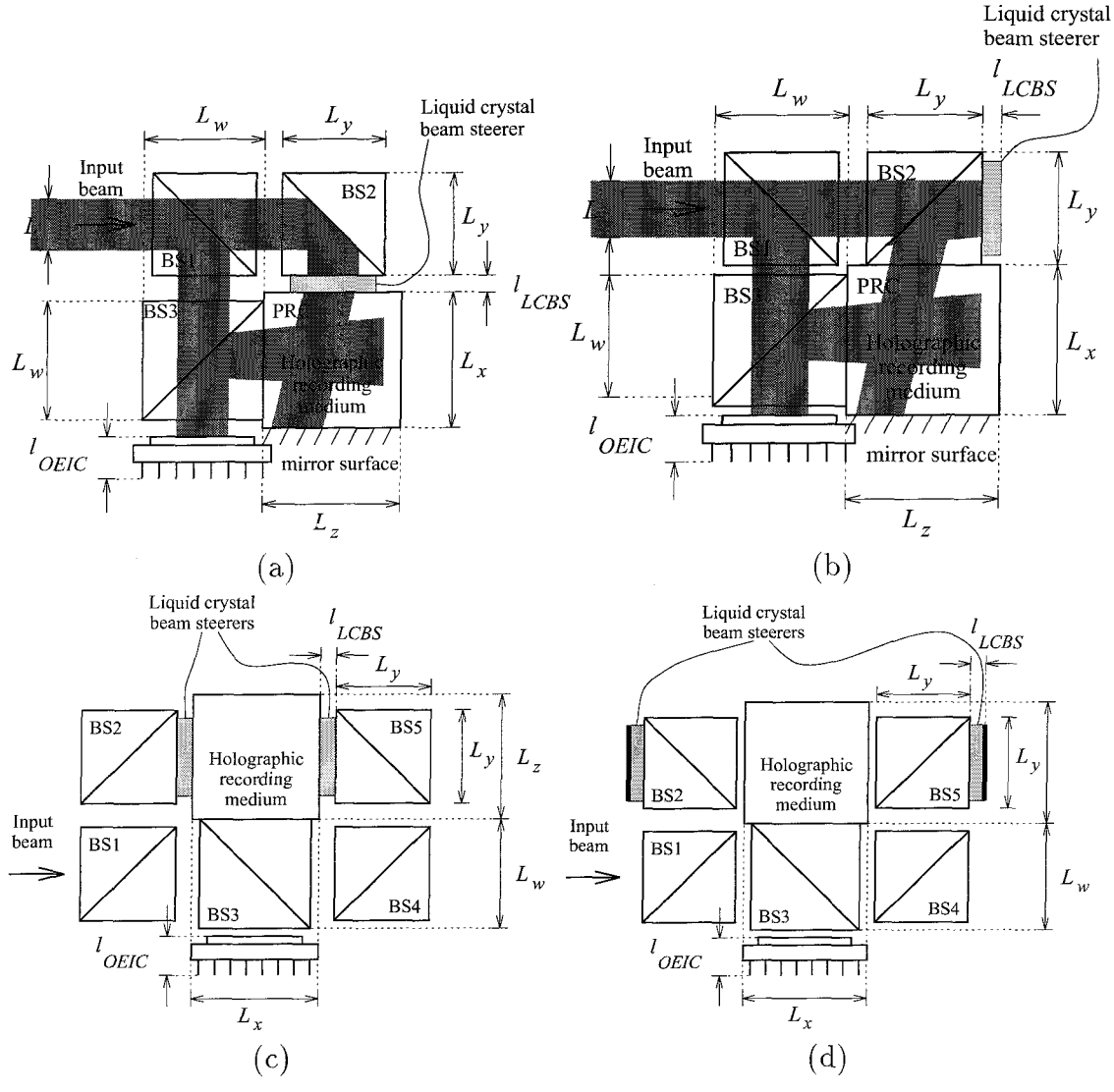


Figure 5.3: Four implementations of the basic compact angularly multiplexed module.

poorer diffraction efficiency and of increased absorption losses (as will be explained in §5.2.3). The modules shown in Figures 5.3c and 5.3d include two beam steerers, and more closely resemble the arrangement illustrated in Figure 1.7.

The four proposed architectures can also be classified according to the type of beam steering unit they employ. While the use of transmissive beam steerers is more straightforward and perhaps less likely to cause undesirable losses, reflective beam steerers are expected to be less expensive, because they can employ silicon substrates processed at low cost in high-volume fabrication facilities. Reflective beam steerers are also potentially faster, since they require only half the thickness of an otherwise equivalent transmissive unit to provide a 2π phase retardation. The switching speed of a nematic liquid crystal modulator is a strong function of the thickness of its liquid crystal film[112]. The modules illustrated in Figures 5.3a and 5.3c include transmissive beam steering units, while those in Figures 5.3b and 5.3d use reflective units.

We now describe how the signal beam and the main and conjugate reference beams are obtained. In each of the four architectures, the input beam is split into two beams by beam splitter BS1. One of these beams is reflected toward the recording medium by BS2, and becomes the main reference beam. The other beam illuminates the OEIC, which reflects it and imprints data on it. The resulting signal beam reaches the recording medium after propagating through BS3. In the modules shown in Figures 5.3c and 5.3d, the counter-propagating reference beam is obtained by routing the input beam around the storage medium using beam splitters BS1, BS3, BS4 and BS5. In the modules illustrated in 5.3a and 5.3b, the main reference is reflected by a metallized face at the bottom of the storage crystal. The reflected beam becomes the reference beam for readout[1]. The operation and consequences of this arrangement will be described in more detail in §5.2.3. The paths of the input, signal and main reference beams are shaded in Figures 5.3a and 5.3b; they have been omitted from the more complex module architectures (Figures 5.3c and 5.3d) for the sake of clarity.

In order to guarantee an efficient use of the limited available input power, the

[1]Note that the beam steering unit must be set to a deflection angle equal in magnitude but opposite to that employed when the data page to be retrieved was recorded.

modules need to be able to route the beams and split the power between them under electronic control. For example, in the modules shown in Figures 5.3c and 5.3d, it is desirable that most of the input power be routed to the conjugate reference during readout operations, to maximize the power diffracted into the photodetectors and the associated signal-to-noise ratio (SNR) (a detailed signal-to-noise ratio analysis of the DHR is provided in §3.3.3). However, in all the modules, the input power must be split between the signal beam and the main reference beam during recording operations, to ensure the fast recording of a strong hologram. Moreover, birefringent elements may be needed to guarantee that the writing beams are vertically polarized, as required for a strong interaction between the signal and reference beams in the 90° geometry. All these requirements are satisfied by using liquid crystal phase retarders or polarization rotators, in conjunction with polarizing beam splitters. These simple liquid crystal devices have been omitted from the diagrams in Figure 5.3 for the sake of simplicity and clarity; however, they have been accounted for in the cost and volume analyses. The modules shown in Figures 5.3a and 5.3b need three phase retarders or polarization rotators, while those in Figures 5.3c and 5.3d require four.

5.2.2 Sizing of the elements

Our density and cost analyses make reasonable assumptions for various material properties, and further assume that a number of system parameters are fixed by technological constraints (see Table 5.2). The wavelength was chosen to ensure the availability of low-cost, high-power laser diodes and of sensitive recording media. $P_{\text{pixel,min}}$ was determined by a noise analysis of a new DHR pixel design, assuming a 1 ms integration time. ζ is the ratio of the angular spacing between holograms, $\Delta\theta$, to the angular selectivity λ/L, where L is the 1-D transverse extent of the input beam. Figure 5.4 illustrates the definition of ϕ, the 1-D fill factor, which relates the size of the square detectors and modulators, $b \times b$, to the pixel pitch, b/ϕ. The pixels are taken to be squares, as is the complete pixel array, whose dimensions are $Nb/\phi \times Nb/\phi$. The 1-D transverse extent of the input beam, L, is set to be just enough to illuminate the

Symbol	Description	Value
n_{BS}	Refractive index of beam splitters	1.52
n_{PRC}	Refractive index of storage medium	2.2
λ	Wavelength in vacuum	670 nm
P_{in}	Input power	120 mW
$(M/\#)^*$	$M/\#$[41] per unit length[a]	2.5 cm^{-1} or 3.42 cm^{-1}
Θ_{max}	Maximum one-sided steerer angular range	6.0°
N_{max}	Maximum 1-D number of pixels[b]	2048
ϕ	1-D fill factor[c]	35%
$P_{\mathrm{pixel,min}}$	Minimum detectable power, per pixel	360 fW
ζ	Dimensionless angular spacing of holograms[d]	2
l_{LCBS}	Thickness of beam steerer	3 mm
l_{OEIC}	Thickness of OEIC	3 mm
l_{LCPR}	Thickness of a polarization rotator	1.4 mm
x_{min}	Minimum feature size of diffractive lens	0.3 μm

[a] $(M/\#)^* = 2.5\mathrm{cm}^{-1}$ in modules using a single beam steerer (Figures 5.3a and 5.3b), and $(M/\#)^* = 3.42\mathrm{cm}^{-1}$ otherwise.

[b] The total maximum number of pixels is $N_{\mathrm{max}} \times N_{\mathrm{max}}$.

[c] See Figure 5.4.

[d] See beginning of §5.2.3.

Table 5.2: Fixed parameters relevant to cost and density analyses.

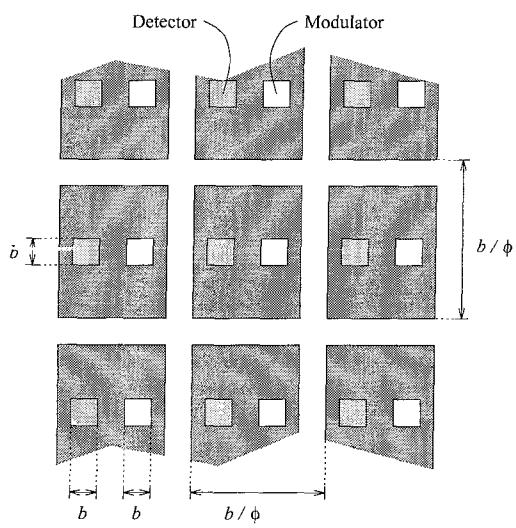


Figure 5.4: Definition of the 1-D fill factor, ϕ.

whole pixel array:

$$L \equiv \frac{Nb}{\phi}. \tag{5.4}$$

Given the pixel array geometry described above and the fixed parameters listed in Table 5.2, the sizes of all the other elements of the storage module are uniquely determined by the modulator/detector size b and the 1-D number of pixels N.

The size $L_w \times L_w \times L_w$ of BS3 is set to be just enough to accommodate the entire central lobe of the signal beam at the exit face. For example, in the modules employing two beam steerers (Figures 5.3c and 5.3d), the signal beam propagates over a folded path of length L_w before reaching the exit face of BS3. The diffraction angle in vacuum is determined by the size of the modulators, b:

$$\theta_s = \frac{\lambda}{b}. \tag{5.5}$$

The size of the beam splitter is thus

$$L_w = L + \underbrace{2\frac{\theta_s}{n_{\mathrm{BS}}}L_w}_{\text{spread}},$$

or

$$L_w = \frac{L}{1 - 2\frac{\theta_s}{n_{\mathrm{BS}}}}. \tag{5.6}$$

In modules using a single beam steerer (Figures 5.3a and 5.3b), the signal beam propagates an additional distance $(L_x - L_w)/2$ between the OEIC and BS3[2]. The following equation expresses the minimum size of BS3:

$$L_w = L + \underbrace{2\frac{\theta_s}{n_{\mathrm{BS}}}L_w}_{\text{spread in BS3}} + \underbrace{\theta_s (L_x - L_w)}_{\text{spread between OEIC and BS3}}$$

[2]The area of the DHR OEIC is larger than that of the pixel array alone. For this reason, we assume that the OEIC extends below other components of the module. In modules employing one beam steerer (Figure 5.3a and 5.3b), the OEIC extends to the right, under the recording medium. Since it is a planar device that must be in a horizontal position, and since $L_w < L_x$, there is a gap between the OEIC and BS3. Because the centers of BS3 and the recording medium are vertically aligned, the dimension of the gap along the signal beam path is $(L_x - L_w)/2$.

$$= \frac{L + \theta_s L_x}{1 - 2\frac{\theta_s}{n_{\mathrm{BS}}} + \theta_s}. \tag{5.7}$$

The 1-D transverse extent L_x of the signal beam at the exit face of the recording medium is determined by its diffraction spread. We set the dimensions[3] of the recording medium to $L_x \times L_z \times L_x$, to accommodate the entire central lobe of the signal beam in the recording material. L_x is simply the size of the signal beam at the exit face of beam splitter BS3, plus the diffraction spread in the recording medium:

$$L_x = L_w + 2\frac{\theta_s}{n_{\mathrm{PRC}}} L_z, \tag{5.8}$$

where L_z is the dimension of the recording medium along the propagation axis of the signal beam. This dimension is chosen to be just large enough to fit the entire width of the reference beam in the recording medium at any angular location:

$$L_z = L + 2rL_y \tan \Theta_{\mathrm{BS}} + 2L_x \tan \Theta_{\mathrm{PRC}}, \tag{5.9}$$

where L_y is the dimension of BS2 along the axes of the module in the plane of Figure 5.3, Θ_{BS} and Θ_{PRC} are the maximum deflection angles of the reference beam in the beam splitters and in the recording medium, respectively, and r is equal to one if reflective beam steerers are employed (Figures 5.3b and 5.3d), and zero otherwise:

$$\Theta_{\mathrm{PRC}} = \arcsin \frac{\sin \Theta}{n_{\mathrm{PRC}}}; \tag{5.10}$$

$$\Theta_{\mathrm{BS}} = \arcsin \frac{\sin \Theta}{n_{\mathrm{BS}}}; \tag{5.11}$$

$$r = \begin{cases} 1 & \text{if beam steerers are reflective (Figures 5.3b and 5.3d)} \\ 0 & \text{otherwise.} \end{cases} \tag{5.12}$$

The dimension L_y of the beams splitters adjacent to the recording medium in the reference path(s), BS2 and BS5, depends upon the type of beam steerers employed by the storage modules. If the beam steerers are transmissive (Figures 5.3a and 5.3c),

[3]Note that the signal beam spreads in two dimensions, one of which is perpendicular to the plane of Figure 5.3.

then the beam splitters need only be large enough to accommodate the width L of the input truncated plane wave. However, BS2 and BS5 must be larger if reflective beam steerers are employed (Figures 5.3b and 5.3d). In these cases, the input beam is first reflected by the beam splitter to the beam steerer, which deflects the beam[4] and reflects it toward the recording medium, through the beam splitter. The total path of the reflected beam on its second pass through the beam splitter is L_y. If we require that the deflected truncated plane wave be entirely contained within the beam splitter, then the minimum dimension[5] of the beam splitter is $L + 2L_y \tan\Theta$. We thus obtain

$$L_y = \frac{L}{1 - 2r\tan\Theta}. \tag{5.13}$$

Substitutions and eliminations can be carried out on equations 5.9, 5.13, 5.8, and equation 5.6 in the case of modules employing two beam steerers (Figures 5.3c and 5.3d) or equation 5.7 in the case of modules with only one beam steerer (Figures 5.3a and 5.3b). We obtain the sets of equations in Tables 5.3 and 5.4, which straightforwardly give the sizes of all the elements, computed in the following order: $L \rightsquigarrow L_y \rightsquigarrow L_z \rightsquigarrow L_x \rightsquigarrow L_w$. The dimensions of the module's optical components are expressed in terms of L_w, L_x, L_y and L_z in Table 5.5.

Reduction of the signal beam spread

In order to faithfully record the information imprinted on the signal beam by the OEIC, the physical support of the holograms must be large enough to encompass the spatial frequencies of the pixel array. This is mathematically expressed by the dependence of equations 5.16 and 5.20 upon θ_s. This lower bound on L_x is detrimental to the storage density achieved by the module. It also contributes to increase the cost of the storage module, since the cost of several individual components is an increasing

[4]The beam steerer deflects the beam by an angle θ such that $-\Theta \leq \theta \leq \Theta$, where Θ is the one-sided angular range of the beam steerer.

[5]We are concerned here with the dimension of the beam splitter along the primary axes of Figure 5.3. The dimension of the beam splitter along the axis perpendicular to Figure 5.3 is simply L, because the beam steerer deflects the input beam in the plane of the Figure.

$$L_z = \frac{L\left(1-2\frac{\theta_s}{n_{\mathrm{BS}}}+2\tan\Theta_{\mathrm{PRC}}\right)+2r\left(1-2\frac{\theta_s}{n_{\mathrm{BS}}}\right)L_y\tan\Theta_{\mathrm{BS}}}{1-2\frac{\theta_s}{n_{\mathrm{BS}}}-4\frac{\theta_s}{n_{\mathrm{PRC}}}\left(1-2\frac{\theta_s}{n_{\mathrm{BS}}}+\theta_s\right)\tan\Theta_{\mathrm{PRC}}} \tag{5.14}$$

$$L_y = \frac{L}{1-2r\tan\Theta_{\mathrm{BS}}} \tag{5.15}$$

$$L_x = 2\frac{\theta_s}{n_{\mathrm{PRC}}}L_z+\frac{L+2\frac{\theta_s^2}{n_{\mathrm{PRC}}}L_z}{1-2\frac{\theta_s}{n_{\mathrm{BS}}}} \tag{5.16}$$

$$L_w = \frac{L+\theta_s L_x}{1-2\frac{\theta_s}{n_{\mathrm{BS}}}+\theta_s} \tag{5.17}$$

Table 5.3: Dimensions of the elements in modules with one beam steerer.

$$L_z = \frac{L\left(1+\frac{2\tan\Theta_{\mathrm{PRC}}}{1-2\frac{\theta_s}{n_{\mathrm{BS}}}}\right)+2rL_y\tan\Theta_{\mathrm{BS}}}{1-4\frac{\theta_s\tan\Theta_{\mathrm{PRC}}}{n_{\mathrm{PRC}}}} \tag{5.18}$$

$$L_y = \frac{L}{1-2r\tan\Theta_{\mathrm{BS}}} \tag{5.19}$$

$$L_x = 2\frac{\theta_s}{n_{\mathrm{PRC}}}L_z+\frac{L}{1-2\frac{\theta_s}{n_{\mathrm{BS}}}} \tag{5.20}$$

$$L_w = \frac{L}{1-2\frac{\theta_s}{n_{\mathrm{BS}}}} \tag{5.21}$$

Table 5.4: Dimensions of the elements in modules with two beam steerer.

Component	Dimensions
Recording medium	$L_x \times L_z \times L_x$
BS1, BS4	$L \times L \times L$
BS2, BS5	$L_y \times L_y \times L$
BS3	$L_w \times L_w \times L_w$

Table 5.5: Dimensions of optical components.

function of their volume[6]. One way to decrease the diffraction spread of the signal beam is to introduce a lens between the OEIC and BS3. This solution may seem a bit ironic and surprising, since we loudly advertised that our integrated storage modules used lens-less data paths and hence eliminated the need for costly high-resolution aberration-corrected lenses and the spacing that is usually required between the lenses and other components. However, we claim that the module suffers none of these drawbacks as a result of the addition of the proposed lens. An inexpensive lens can be employed, because phase-conjugate readout undoes linear phase aberrations. For the same reason, the position of the lens is not critical, and the module's components can still be abutted and assembled in a compactly integrated package. In particular, a diffractive lens can be fabricated on the surface of the OEIC, or on the surface of BS3, in which cases its addition entails no volume increase.

The main effect of the lens is to decrease the diffraction spread of the signal beam. Equations 5.14–5.21 remain valid; however, the following is substituted for equation 5.5:

$$\theta_s = \begin{cases} \frac{\lambda}{b} - \frac{L}{2F} = \frac{\lambda}{b} - \frac{1}{2\sqrt{2}(F/\#)} & \text{if } (F/\#) \geq \frac{b}{2\sqrt{2}\,\lambda} \\ 0 & \text{otherwise,} \end{cases} \tag{5.22}$$

where F is the focal length of the lens, and $F/\# \equiv F/(\sqrt{2}\,L)$ is its *F-number*[7]. The condition

$$(F/\#) = \frac{b}{2\sqrt{2}\,\lambda} \tag{5.23}$$

corresponds to the situation where the spectral components associated with the highest spatial frequency $1/b$ in the passband of the system propagate horizontally due to the focusing action of the lens; in other words, the diffraction spread is exactly eliminated, and the cross-section of the portion of the signal beam representing the desired information is a square measuring L on a side. If a faster lens is employed, the minimum size of the optical elements (e.g., BS3) remains L, the size of the pixel array.

From equation 5.22, we expect the storage density of the module to improve as

[6]The cost models will be explained in §5.2.5.

[7]The minimum aperture of the lens is a square measuring L on a side, since this is the dimension of the pixel array. The diameter of the smallest circle in which the lens can be inscribed is $\sqrt{2}\,L$.

the $F/\#$ is decreased, until equation 5.23 is satisfied. This condition may not always be satisfied, however, because technological limitations may prevent the economical fabrication of lenses with the desired properties. We will illustrate this issue in the context where a diffractive lens is employed. A number of circular patterns[8] encode the desired quadratic phase profile on the surface of the diffractive lens. The width of the fringes corresponding to positive and negative phase shrinks as we move away from the center of the lens. The radial extent of a fringe where the phase keeps the same polarity is approximately $\lambda F/(2r)$, where r is the radius of the circular fringe. In order for the lens to possess the desired focusing properties, this fringe width should be a few times larger than the minimum feature size $x_{\min}$ of the lithography system used to fabricate the lens. In what follows, we will, somewhat arbitrarily, assume that the minimum fringe width is $5x_{\min}$. Based on the considerations above, the minimum achievable $F/\#$ (i.e., the maximum focusing power) is given by

$$(F/\#)_{\min} = 5\frac{x_{\min}}{\lambda}. \tag{5.24}$$

We will see in §5.2.6 that the presence of a simple diffractive lens in the signal path can considerably increase the storage density of the module, and also reduce its cost.

5.2.3 Number of holograms

The number M of holograms, or data pages, that can be multiplexed in the recording medium is a function of the angular range of the beam steerer, of the angular selectivity of the medium, and of the available laser power. The angular selectivity, defined as the angular spacing between the nulls of the diffraction efficiency of a single plane-wave hologram, is λ/L in the 90° geometry, where L is the transverse extent of the beams, and also the effective interaction length. The angular separation θ_{mux} between holograms is set to be a multiple of the angular selectivity:

$$\theta_{\text{mux}} = \frac{\zeta\lambda}{L}, \zeta \in \{1, 2, 3, ...\}. \tag{5.25}$$

[8]The circular fringes may be truncated in the corners of the square lens.

The highest hologram packing density is obtained by setting $\zeta = 1$; however, this could result in intolerably high cross talk[113] in high-bandwidth images, since the sidelobes adjacent to the first null are strong. For this reason, in what follows, we choose

$$\zeta = 2, \tag{5.26}$$

in agreement with Table 5.2.

If we assume that the number of multiplexed holograms is not limited by power considerations, then it is simply given by the ratio of the total angular range of the beam steerer[9] to the angular separation between holograms, θ_{mux}, given by equation 5.25:

$$M = \left\lfloor \frac{2\Theta L}{\zeta\lambda} \right\rfloor. \tag{5.27}$$

The power diffracted into the photodetector associated with each pixel is

$$P_{\text{pixel}} = P_{\text{ref}} \left(\frac{\xi(M/\#)}{MN} \right)^2 = P_{\text{ref}} \left(\frac{\xi\zeta(M/\#)^*\lambda}{2\Theta N} \right)^2,$$

where ξ is a unit-less measure of the worst-case overlap between the reference beam for readout and the volume in which the hologram is stored. ξ depends on the module architecture, on the dimensions of the components, and on the angular range of the beam steerer; mathematical expressions for this overlap factor will be derived in a moment.

In order to ensure the accurate detection of the optical data page, the diffracted power per pixel must be at least $P_{\text{pixel,min}}$ (see Table 5.2):

$$P_{\text{ref}} \left(\frac{\xi(M/\#)}{MN} \right)^2 \geq P_{\text{pixel,min}}. \tag{5.28}$$

If this condition is not satisfied with the number of holograms M obtained from equation 5.27, then M must be reduced to the maximum integer value that satisfies

[9] We recall that Θ is defined as the *one-sided* angular range of the beam steerer.

it. In summary, the number of multiplexed holograms is

$$M = \begin{cases} \left\lfloor \frac{2\Theta L}{\zeta\lambda} \right\rfloor & \text{if } P_{\text{ref}} \left(\frac{\xi\zeta (M/\#)^* \lambda}{2\Theta N} \right)^2 \geq P_{\text{pixel,min}} \\ \left\lfloor \frac{L\xi (M/\#)^*}{N} \sqrt{\frac{P_{\text{ref}}}{P_{\text{pixel,min}}}} \right\rfloor & \text{otherwise.} \end{cases} \tag{5.29}$$

In the equations above, P_{ref} is the total power of the reference beam for readout (the conjugate reference). In modules employing two transmissive beam steerers (Figure 5.3c), $P_{\text{ref}} \approx P_{\text{in}}$, since it is assumed that almost all the input power can be routed to the reference beam for readout by the combined action of polarizing beam splitters and liquid crystal devices. However, in the modules employing a single beam steerer (Figures 5.3a and 5.3b), the main reference beam first traverses the recording medium before it is reflected to yield the desired conjugate reference; the power available for readout is thus decreased by the absorption of the recording medium. This absorption can be substantial. In photorefractive materials, it is usually optimized to provide enough recording sensitivity while ensuring that enough power is available to produce a strong and reasonably uniform holographic reconstruction; the optimum power transmittance is around $\exp(-1)$. Finally, in modules employing reflective beam steerers (Figures 5.3b and 5.3d), additional losses are incurred because of the beam splitters BS2 and BS5. The motivation for employing reflective beam steerers is that such devices can employ conventionally processed silicon die as backplanes, and hence leverage the low cost of silicon fabrication. While there are ways to eliminate the beam splitter losses[10], none of those we imagined were compatible with an inexpensive conventionally fabricated VLSI steerer backplane. In the simple case where a 50% beam splitter is employed, 50% of the input light power is lost in each pass.

The two passes of the beam through the recording medium in modules using a single beam steerer (Figures 5.3a and 5.3b) have an additional undesirable effect:

[10]For example, one could use a polarizing beam splitter and insert a quarter-wave plate with its principal axes at 45° with respect to those of the beam steerer between the electro-optic layer and the reflectors within the beam steerer.

holograms are erased faster. We recall the definition of $(M/\#)^*$,

$$(M/\#)^* \equiv \frac{\partial}{\partial z}\frac{A_0\tau_e}{\tau_w},$$

where A_0^2 is the saturation diffraction efficiency, τ_w and τ_e are respectively the recording and erasure time constants, and the z axis is parallel to the propagation axis of the signal beam. The faster erasure can be accounted for by decreasing the effective $(M/\#)^*$, as indicated in Table 5.2.

The value of $(M/\#)^*$ was set to 3.4/cm for modules employing two beam steerers, and to 2.5/cm for those using a single beam steerer. These values may seem somewhat low compared to the results of measurements mentioned in §1.2.3. However, these measurements were carried out at relatively short wavelengths, with green or blue light, to which popular photorefractive recording materials are the most sensitive. However, no inexpensive and sufficiently powerful laser sources are currently commercially available at those wavelengths; virtually all laser diodes operate in the red or infrared. Cost considerations therefore dictate a compromise between the sensitivity range of recording media and the operating wavelength range of powerful, inexpensive laser diodes. We chose to employ a 670 nm (red) laser diode and to accept suboptimal $(M/\#)^*$ values.

The development of laser diodes operating at shorter wavelengths and of materials sensitive at longer wavelength may bridge the "wavelength gap" over the coming years. Such developments would result in an increased $(M/\#)^*$. The effects of a hypothetical 45% increase of $(M/\#)^*$ are summarized in Tables 5.13 and 5.14 at the end of §5.2.6.

Overlap between readout reference and storage volume

We now turn to the problem of calculating the reduction of the diffraction efficiency caused by the shift between the main and conjugate reference beams. The origin of this shift is the fact that the pivot points around which the reference beams (main and conjugate) are steered are located in the center of the planes of the beam steerers, instead of the center of the recording material. The shift amount is different in the various

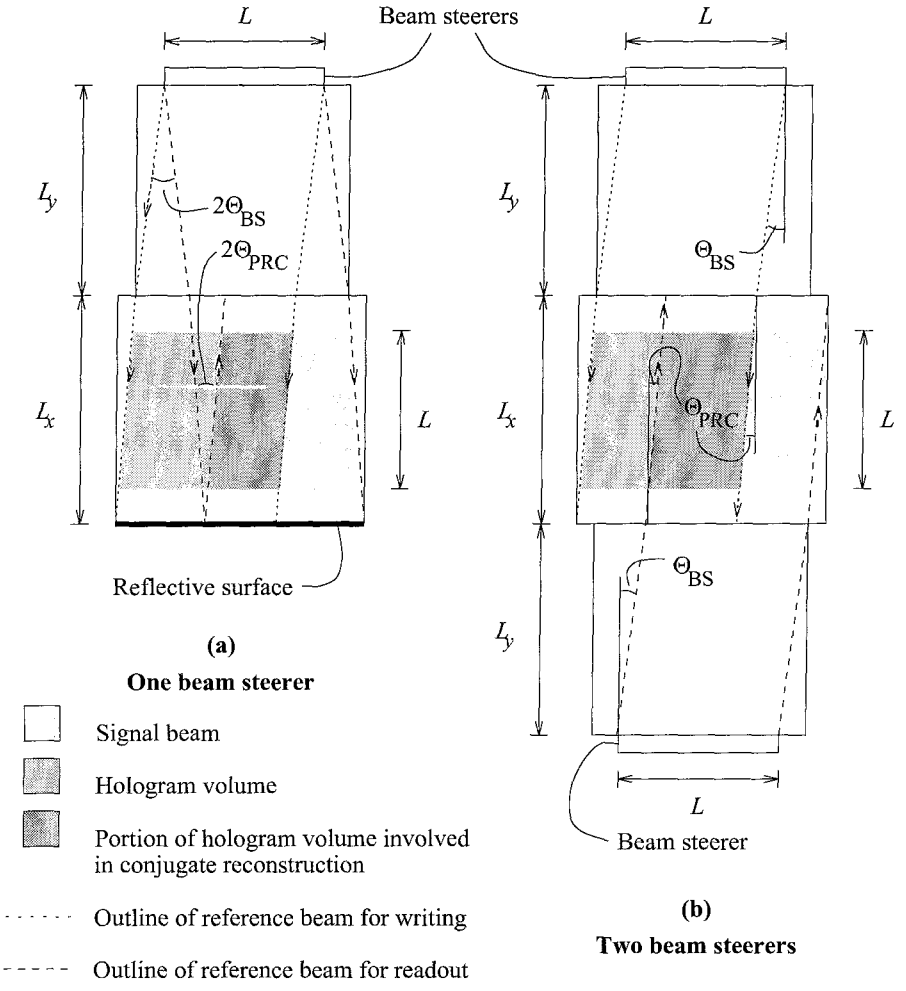


Figure 5.5: Shift of the conjugate reference beam with respect to the main reference beam.

module architectures. For example, Figure 5.5a shows the geometry relevant to the case of a module employing a single reflective beam steerer (see also Figure 5.3b). At the top of the illustration is an unfolded representation of BS2; the recording medium is in the middle. The worst overlap (weakest diffraction efficiency) occurs at the maximum deflection angle, Θ. The corresponding angles inside the beam splitter and the recording material are Θ_{BS} and Θ_{PRC}, respectively. Let's say that the hologram is recorded while the beam steerer deflects the reference toward the left by an angle Θ. The hologram is stored in the volume of the recording material defined by the intersection of the reference and signal beams. To read this data page out, the beam steerer is pro-

grammed to deflect the reference beam by an angle Θ *toward the right*, so that, after it traverses the recording medium and is reflected by the surface at the bottom of the medium, the beam is a phase-conjugate version of the reference beam that wrote the hologram. Note that only a slice of the hologram is illuminated by the conjugate reference beam. Neglecting weak high-order re-diffraction effects, only this fraction of the hologram's volume contributes to the reconstruction. The overlap factor ξ introduced in §5.2.3 is defined as this fraction. The diffracted field is proportional to ξ, and the diffracted power is proportional to ξ^2. The horizontal distance between the left edge of the hologram volume and the conjugate reference is $2L_y \tan\Theta_{\mathrm{BS}} + 2L_x \tan\Theta_{\mathrm{PRC}}$, and[11] $\xi = \max\left(0, \left(L - 2L_y \tan\Theta_{\mathrm{BS}} - 2L_x \tan\Theta_{\mathrm{PRC}}\right)/L\right)$. If a transmissive beam steerer is substituted for the reflective unit (Figure 5.3a), the lateral translation in the beam splitter is eliminated, and $\xi = \max\left(0, \left(L - 2L_x \tan\Theta_{\mathrm{PRC}}\right)/L\right)$.

The situation where two reflective beam steerers are used (Figure 5.3d) is illustrated in Figure 5.5b. In this case, $\xi = \max\left(0, \left(L - 2rL_y \tan\Theta_{\mathrm{BS}} - L_x \tan\Theta_{\mathrm{PRC}}\right)/L\right)$. In summary,

$$\xi = \max\left(0, \frac{L - 2rL_y \tan\Theta_{\mathrm{BS}} - (3 - N_{\mathrm{LCBS}})\, L_x \tan\Theta_{\mathrm{PRC}}}{L}\right), \tag{5.30}$$

where N_{LCBS} is the number of beam steerers per module.

Angular range of beam steerer

In the analysis leading to equation 5.29, we saw that if

$$P_{\mathrm{ref}} \left(\frac{\xi\zeta(M/\#)^*\lambda}{2\Theta N}\right)^2 < P_{\mathrm{pixel,min}},$$

the number of holograms is limited by power considerations, and the angular range of the beam steerer is not entirely used. A large angular range can be too much of a good thing, since it is detrimental to the module's storage density (the dimensions of its elements depend on the one-sided angular range Θ through equations 5.14–5.21).

[11] The function max is defined as its maximum argument.

In cases where the number of holograms is limited by diffracted power considerations rather than the maximum angular range $\Theta_{\max}$ provided by the chosen beam steerer technology, the angular range should be set to the minimum value that satisfies

$$P_{\text{ref}}\left(\frac{\xi\zeta(M/\#)^*\lambda}{2\Theta N}\right)^2 \leq P_{\text{pixel,min}}:$$

$$\Theta = \begin{cases} \Theta_{\max} & \text{if } P_{\text{ref}}\left(\frac{\xi\zeta(M/\#)^*\lambda}{2\Theta_{\max}N}\right)^2 \geq P_{\text{pixel,min}} \\ \max\left\{\Theta : P_{\text{ref}}\left(\frac{\xi\zeta(M/\#)^*\lambda}{2\Theta N}\right)^2 \geq P_{\text{pixel,min}}\right\} & \text{otherwise.} \end{cases} \tag{5.31}$$

5.2.4 Storage density

The raw storage density of the module is the total number of stored bits, MN^2, divided by the volume V of the module. V is defined as the volume of the smallest rectangular parallelepiped that contains all the elements of the module, assembled in their functional topology. V depends upon the architecture of the module:

$$V = \begin{cases} L_x\left(\max\left(\frac{L_w}{2}, \frac{L}{2} + l_{\text{LCPR}}\right) + \frac{L_w}{2} + L_z\right)(L_x + L_y + l_{\text{OEIC}} + l_{\text{LCBS}}) & \text{Fig. 5.3a} \\ L_x\left(\max\left(\frac{L_w}{2}, \frac{L}{2} + l_{\text{LCPR}}\right) + \frac{L_w}{2} + \frac{L_z}{2} + \max\left(\frac{L_z}{2}, \frac{L_y}{2} + l_{\text{LCBS}}\right)\right) & \text{Fig. 5.3b} \\ \quad \times (L_x + L_y + l_{\text{OEIC}}) & \\ L_x\left(L_w + 10\ \mu\text{m} + L_z + l_{\text{OEIC}}\right)(L_x + 2L_y + 2l_{\text{LCBS}} + l_{\text{LCPR}}) & \text{Fig. 5.3c} \\ L_x\left(L_w + 10\ \mu\text{m} + L_z + l_{\text{OEIC}}\right)(L_x + 2L_y + 2l_{\text{LCBS}})\,. & \text{Fig. 5.3d} \end{cases} \tag{5.32}$$

The storage density, in bits/m^3, is simply

$$\mathcal{D} = \frac{MN^2}{V}, \tag{5.33}$$

where M is obtained from equation 5.29 and V is from equation 5.32 above. Numerical results based on the analysis above are presented in §5.2.6.

5.2.5 Cost

The storage density, for which a model was developed in §5.2.4, is one of the main properties of a data storage technology. Another important property is its cost per amount of data stored, often expressed in dollars per megabyte ($/MB). We now present a model for estimating the cost of integrated modular holographic data storage systems. More specifically, we will attempt to evaluate the cost of such systems in high volume production, using cost models and performance projections for the various components for mid- to late-1998. This is admittedly a subject prone to controversy, and the validity of the models may be cast in doubt by unexpected technological advances (or lack thereof) or market conditions. A good cost model is nonetheless of the utmost importance in determining the best target markets for the new technology and in assessing its competitiveness.

We tried to firmly root our cost models in realistic projections for the costs of its various components. All the component cost models are based on high-volume production; non-recurring engineering (NRE) costs are therefore neglected. The issue of how integrated modular holographic memories would reach the assumed high sales volume (e.g., initial sales in a low-volume, highly lucrative market, or a large upfront investment) is not addressed. However, most components can leverage investments made for other markets. For example, the DHR OEIC employs a conventionally processed CMOS VLSI die as a backplane, and LCOS technology is being targeted for consumer electronics and projection display markets. Beam steerers have applications as routing switches in fiber optic communication systems. Some photorefractive crystals such as lithium niobate are produced in high volumes as surface acoustic wave (SAW) filters for wireless communication applications.

Note that we do not claim that integrated modular holographic memory will reach high-volume production by the end of 1998; we merely try to assess its potential,

Symbol	Description	Cost
$C_{\text{a,Si}}$	Areal cost of silicon (DHR OEIC)	\$15/cm^2
$C_{\text{a,PRC}}$	Areal cost of photorefractive crystal	\$0.83/cm^2
$C_{\text{v,PRC}}$	Cost of photorefractive crystal per volume	\$5/cm^3
$C_{\text{a,rLCBS}}$	Areal cost of reflective beam steerer	\$20/cm^2
$C_{\text{a,tLCBS}}$	Areal cost of transmissive beam steerer	\$50/cm^2
C_{laser}	Cost of 120 mW diode laser	\$12
$C_{\text{collimation}}$	Cost of collimating a laser beam	\$2
C_{LCPR}	Cost of a simple polarization rotator	\$0.75
C_{LCOS}	Cost of a LCOS modulator[a]	\$5
C_{assembly}	Module assembly cost, Fig. 5.3a or 5.3b	\$4.00
	Module assembly cost, Fig. 5.3c or 5.3d	\$4.50
$C_{\text{signal lens}}$	Cost of lens in signal beam[b]	\$2.00

[a]This is only the cost of the liquid crystal modulator structure; it does not include the cost of the silicon backplane.

[b]See section 5.2.2.

Table 5.6: Parameters of the cost model.

unhindered by NRE costs that will eventually become insignificant if the technology is successful.

Table 5.6 shows the parameters of the cost model. The cost of many components is a function of their dimensions.

Optoelectronic integrated circuit

The cost of the OEIC is comprised of the cost of the silicon backplane and of the cost of its liquid crystal modulators:

$$C_{\text{OEIC}} = C_{\text{a,Si}} A_{\text{OEIC}} + C_{\text{LCOS}}, \tag{5.34}$$

where A_{OEIC} is the area of the silicon die, comprised of the pixel array and of support functions (electronic page buffer, signal processing circuits, control functions, pads). Together, the support functions are assumed to amount to a 125% overhead[12] over

[12]This overhead does not necessarily increase the volume of the module, as defined in §5.2.4, since the optically inactive parts of the die can be located under module components horizontally adjacent to BS3 (see Figure 5.3).

the area of the pixel array. The total silicon area is thus

$$A_{\mathrm{OEIC}} = 2.25L^2. \tag{5.35}$$

The maximum area of the die $A_{\mathrm{OEIC},max}$ is taken to be 300 mm^2, which is the size limit forecast by the Semiconductor Industry Association's National Technology Roadmap for Semiconductors[13] (NTRS) for high-volume integrated circuits in 1998[111, Table 1B]. The number of pixels $N \times N$ is limited to 2048×2048, which is approximately the maximum number of pixels in devices under development or planned for the near future. Any solution leading to a larger area or a larger number of pixels is rejected from our analysis.

The areal cost of the DHR silicon, $C_{\mathrm{a,Si}}$, was chosen as an intermediate areal cost between the Semiconductor Industry Association (SIA) 1998 projections for DRAM and microprocessors[111, Tables IA and IB], to reflect the intermediate complexity of the fabrication process of this device, as well as its production volume. The SIA projections are consistent with areal costs of \$6.71/cm^2 and \$35/cm^2 for DRAM and high-density logic (e.g., microprocessors), respectively[111, Tables IA and IB, again].

Beam steerers

We assume that programmable liquid crystal optical phased arrays similar to those developed by Raytheon Electronic Systems[67, 68, 69] are employed as addressing devices in our modules. These devices are based on an array of striped electrodes that control the phase profile of an overlying layer of liquid crystal. The electrodes are patterned lithographically, using a process similar to that employed to pattern metal interconnections in integrated circuits. The cost model of the beam steerers is therefore similar to that of the DHR silicon. The areal cost of reflective beam steerers is set to \$20/cm^2, or 33% above that of the DHR, but more than 40% below that of microprocessors. The number of masks involved is less than in the case of the DHR OEIC; however, the lithography requirements may be more stringent, and the cost of

[13]This document was formerly known as the *SIA Roadmap*.

the modulating layer must be accounted for.

Transmissive beam steerers are assumed to incur a 150% cost overhead[14] compared to reflective beam steerers with similar properties. This overhead accounts for two facts. First, transmissive beam steerers cannot use the same fabrication process as their reflective counterparts, which benefit from low-cost silicon substrate processing. While the processing techniques are related, transmissive beam steerers must rely on transparent electrodes on a transparent substrate (e.g., glass) or on sophisticated silicon processing (e.g., the epitaxial lift-off technique developed by Kopin Corporation (Taunton, MA)). Therefore, transmissive beam steerers cannot leverage silicon technology as effectively as their reflective counterparts. Second, for the foreseeable future, transmissive beam steerers are not expected to attain the high fabrication volumes common in the RAM market and other semiconductor markets. The reader familiar with silicon pricing[15] will recognize that sales volumes can account for threefold cost differences for devices using the same silicon area and the same process, even if the devices sell in volumes of thousands or more.

Another motivation for employing reflective beam steerers is their higher speed. All other things being equal, a transmissive beam steerer requires twice the liquid crystal layer thickness of a reflective unit, because of the two passes of the beam through the liquid crystal in a reflective unit. The switching time of a nematic liquid crystal goes roughly as the square of the cell thickness[112, 97]. Therefore, a reflective beam steerer can be approximately four times faster than its transmissive counterpart.

However, as we will see in §5.2.6, the advantages of reflective beam steerers pale in comparison with the optical power losses encountered in practical systems employing them.

[14]In other words, the cost of a transmissive beam steerer is 2.5 times that of a reflective unit with otherwise comparable properties.

[15]Silicon devices are fabricated using processes related to those used in transmissive beam steerers, although the latter are more expensive. We can reasonably expect similar economies of scale.

Photorefractive crystal

Our cost model for the recording medium is comprised of two components: an areal cost, which accounts for surface treatments such as polishing and anti-reflection coatings, and a volume cost, which represents growth cost. The relative importance of these two components varies from one material to another. For example, the growth of $LiNbO_3$ is inexpensive; most of the cost of a $LiNbO_3$ holographic crystal represents surface treatments. At the other end of the spectrum, it is currently difficult to grow sizeable $BaTiO_3$ crystals suitable for photorefractive storage and possessing good optical quality. The cost of $BaTiO_3$ crystals is thus dominated by volume cost. Progress is being made in this area, however. For instance, Dr. Demetri Psaltis recently reported seeing a high-quality two-inch $BaTiO_3$ boule grown in China.

Given the significant ongoing developments in the field of inorganic holographic crystals, we decided not to choose any particular material for our analysis. Instead, we simply assume that

- the cost of a 1 cm^3 cubic crystal is equally split between areal and volume costs;
- the cost of this 1 cm^3 cubic crystal of an as-yet unspecified material is equal to that of a similarly sized $LiNbO_3$:Fe crystal.

The cost of a 1 cm^3 cubic $LiNbO_3$:Fe crystal is estimated at \$10 in high production volume, according to crystal grower Deltronics. We obtain the areal and volume costs shown in Table 5.6 based on this estimate and on the assumptions above. The cost model for the recording medium is thus

$$C_{\mathrm{PRC}} = C_{\mathrm{a,PRC}} A_{\mathrm{PRC}} + C_{\mathrm{v,PRC}} V_{\mathrm{PRC}}, \tag{5.36}$$

where

$$\begin{aligned} A_{\mathrm{PRC}} &= 4L_x L_z + 2L_x^2, \\ V_{\mathrm{PRC}} &= L_x^2 L_z, \\ C_{\mathrm{a,PRC}} &= \$0.8333/\mathrm{cm}^2, \end{aligned}$$

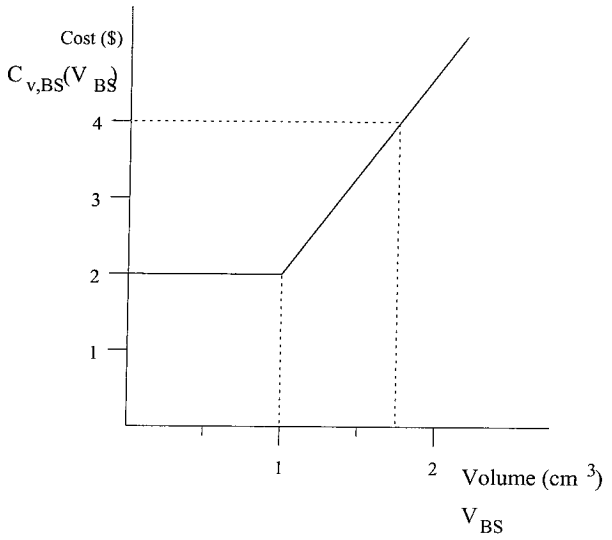


Figure 5.6: Volume cost of beam splitters.

$$C_{\text{v,PRC}} = \$5.00/\text{cm}^3.$$

Beam splitters

The cost model for cubic beam splitters such as those shown in Figure 5.3 (polarizing or not) is based in part on information graciously provided by Dr. Fai Mok of Holoplex. As mentioned in §1.2.5, the advent of mass-marketed optoelectronic information storage and processing devices such as CD readers has driven down the price of optical components such as beam splitters. The cost of beam splitters is a non-decreasing function of their volume. A 1 cm^3 cubic beam splitter can cost as little as \$2, while a 1.75 cm^3 unit costs approximately \$4. We choose a piecewise linear model, illustrated in Figure 5.6, and denote the cost function $C_{\text{v,BS}}(V_{\text{BS}})$, where V_{BS} is the volume of a beam splitter. The volume of each beam splitter is obtained from Table 5.5 and equations 5.14–5.21. While the cost would arguably increase faster than linearly past a certain volume, we will see in §5.2.6 that the volume of the beam splitters never exceeds a few cm^3.

The cost model for beam splitters is hence as follows (see Figure 5.3):

$$C_{\text{BS}} = \begin{cases} C_{\text{v,BS}}\left(L^3\right) + C_{\text{v,BS}}\left(LL_y^2\right) + C_{\text{v,BS}}\left(L_w^3\right) & \text{Fig. 5.3a and 5.3b} \\ 2\,C_{\text{v,BS}}\left(L^3\right) + 2\,C_{\text{v,BS}}\left(LL_y^2\right) + C_{\text{v,BS}}\left(L_w^3\right) & \text{Fig. 5.3c and 5.3d.} \end{cases} \tag{5.37}$$

Equation 5.37 accounts for all beam splitters in a module.

Laser

We assume that an inexpensive ($\approx$ \$12) 670 nm laser diode provides the single input beam required by each of the modules illustrated in Figure 5.3. The model allows for laser sharing between modules, provided that enough power is available to each in order to enable parallel module operation. More specifically, the number of modules sharing a single laser is

$$N_{\text{modules per laser}} = \begin{cases} \left\lfloor \frac{P_{\text{ref}}}{P_{\text{pixel,min}}} \left(\frac{\xi\zeta(M/\#)^*\lambda}{2\Theta N}\right)^2 \right\rfloor & \text{if } P_{\text{ref}} \left(\frac{\xi\zeta(M/\#)^*\lambda}{2\Theta N}\right)^2 \geq P_{\text{pixel,min}} \\ 1 & \text{otherwise,} \end{cases} \tag{5.38}$$

where P_{ref} must take into account the absorption of the recording medium in the modules illustrated in Figures 5.3a and 5.3b, and the losses in BS2 and BS5 in those illustrated in Figures 5.3b and 5.3d.

5.2.6 Numerical results

Given the parameters listed in Tables 5.2 and 5.6, the dimensions of all the elements in a module, and ultimately the storage density and cost of the module, are entirely determined by the modulator/detector size b and the number of pixels in a data page, $N \times N$. We have implemented the density and cost analyses presented above on a computer and have optimized the density and cost[16] over the following ranges:

$$2.75\ \mu\text{m} \leq b \leq 10\ \mu\text{m}$$

[16]The calculations were performed by Mathematica 2.2 running on a Sun SparcStation 20 with two CPUs.

Part of Fig. 5.3	Lens[a]	Max. density[b]	Pixel size[c]	N	Cost at max. dens.
a	No	1.1×10^{14}	7.9 μm	472	\$3.01/MB
a	Yes	2.1×10^{14}	7.9 μm	504	\$2.61/MB
b	No	5.4×10^{13}	7.9 μm	328	\$10.49/MB
b	Yes	1.0×10^{14}	7.9 μm	344	\$9.56/MB
c	No	1.1×10^{14}	7.9 μm	1152	\$0.76/MB
c	Yes	2.0×10^{14}	7.9 μm	1184	\$0.69/MB
d	No	7.0×10^{13}	7.9 μm	488	\$3.01/MB
d	Yes	1.3×10^{14}	7.9 μm	504	\$2.82/MB

[a]In the signal beam.
[b]In bits/m^3.
[c]This is b/ϕ.

Table 5.7: Results of density optimization.

$$64 \leq N \leq 2048.$$

The minimum modulator/detector size b was chosen to be slightly more than four times larger than a wavelength and significantly larger than cell gaps used with high-birefringence liquid crystals. Tables 5.7 and 5.8 summarize the results of this optimization process.

The module architectures that provide the densest data storage are those illustrated in Figures 5.3a and 5.3c. The highest raw density with a diffractive lens present in the signal beam path is 2.1×10^{14} bits/m^3, obtained with the module shown in Figure 5.3a; without such a lens, it drops to 1.1×10^{14} bits/m^3, and the two modules provides approximately the same raw density[17], although that shown in Figure 5.3c offers a fourfold improvement in storage cost. The storage costs associated with the optimized densities are \$2.61/MB (Figure 5.3a, lens present), \$3.01/MB (Figure 5.3a, no lens) and \$0.76/MB (Figure 5.3c, no lens).

The lowest cost per MB is obtained with the module architecture shown in Figure 5.3c: \$0.64/MB with a lens in the signal beam path, \$0.70/MB otherwise; the corresponding raw densities are 1.8×10^{14} bits/m^3 and 9.5×10^{13} bits/m^3, respectively.

[17]The module illustrated in Figure 5.3c has a very slight edge.

Part of Fig. 5.3	Lens[a]	Min. cost	Pixel size[c]	N	Dens.[b] at min. cost
a	No	\$1.14/MB	7.9 μm	1432	5.9×10^{13}
a	Yes	\$0.99/MB	7.9 μm	1456	1.2×10^{14}
b	No	\$1.77/MB	7.9 μm	1440	2.9×10^{13}
b	Yes	\$1.52/MB	7.9 μm	1448	6.1×10^{13}
c	No	\$0.70/MB	7.9 μm	1440	9.5×10^{13}
c	Yes	\$0.64/MB	7.9 μm	1464	1.8×10^{14}
d	No	\$0.99/MB	7.9 μm	1432	4.7×10^{13}
d	Yes	\$0.88/MB	7.9 μm	1448	8.7×10^{13}

[a] In the signal beam.
[b] In bits/m^3.
[c] This is b/ϕ.

Table 5.8: Results of cost optimization.

The properties of these two modules are listed in Table 5.9; their cost breakdown is shown is Table 5.10.

In what follows, we mainly focus our attention on the module illustrated in Figure 5.3c (with or without a lens), as opposed to that shown in Figure 5.3a, for the following reasons:

- The optimized storage cost of the module of Figure 5.3c is more than 35% lower than that of the module in Figure 5.3a (with or without a lens);
- The optimized raw storage density of the module in Figure 5.3c is within 5% of that of the module in Figure 5.3a;
- When the modules are optimized for cost, the modules illustrated in Figure 5.3c possesses a better storage density than that of Figure 5.3a.

The modules with the lowest storage cost share the following properties:

Pixel size The pixels have the minimum allowed size.

Size of OEIC The size of the OEIC is very close to the maximum allowed dimension.

Symbol	Description	Without lens	With lens
	Cost per raw megabyte	\$0.698	\$0.635
b	Modulator/detector size[a]	2.75 μm	2.75 μm
	Pixel pitch	7.86 μm	7.86 μm
M	Number of holograms	1444	1468
N	1-D number of pixels per page	1440	1464
Θ	Angular range	4.9°	4.9°
	Raw module capacity (bits)	2.99×10^9	3.15×10^9
$F/\#$	F-number of lens	N/A	2.24
L	1-D dimension of pixel array	11.3 mm	11.5 mm
L_w	Dim. of BS3	16.7 mm	13.0 mm
L_x	Max. transverse dim. of signal	19.5 mm	13.9 mm
L_y	Dim. of reference beam splitters	11.3 mm	11.5 mm
L_z	Dim. of recording medium[b]	12.8 mm	12.6 mm
V	Volume of module	31.4 cm^3	17.68 cm^3
	Raw storage density[c]	9.54×10^{13}	1.78×10^{14}
A_{OEIC}	Silicon area of DHR OEIC	2.88 cm^2	2.98 cm^2
	Modules per laser	1	1

[a] See Figure 5.4.
[b] Along the propagation axis of the signal beam.
[c] In bits/m^3.

Table 5.9: Properties of modules with lowest storage cost (Figure 5.3c).

Description	Cost without lens	Cost with lens
Recording medium	\$39.04	\$21.33
DHR OEIC	\$48.20	\$49.66
Beam steerers (2)	\$128.01	\$132.32
Beam splitters	\$12.48	\$11.50
Liquid crystal polarization rotators	\$3.00	\$3.00
Source (laser & collimator)	\$14.00	\$14.00
Diffractive lens	\$0.00	\$2.00
Assembly	\$4.50	\$4.50
Total	**\$249.25**	**\$238.30**

Table 5.10: Cost breakdown of modules with lowest storage cost (Figure 5.3c).

Number of holograms and angular range of beam steerer The number of holograms is limited by the available power and by the noise floor. The angular range Θ is 25% less than $\Theta_{\max}$.

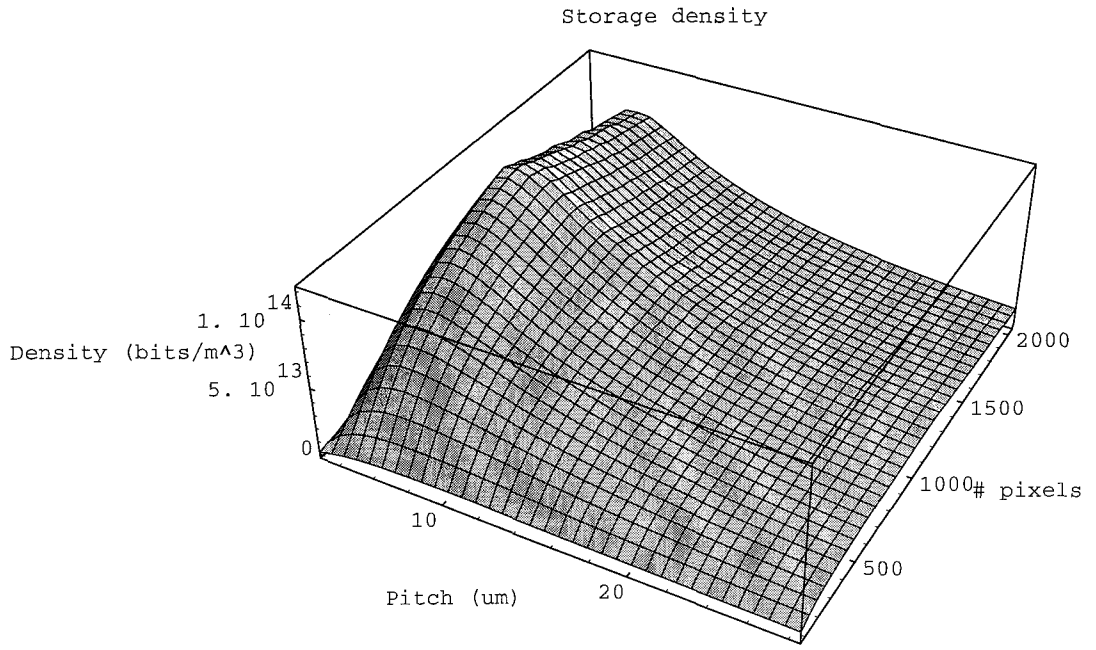
As expected, the addition of a diffractive lens in the signal beam path significantly reduces the volume of the module, by eliminating the diffraction spread of the signal beam, thereby shrinking the dimensions L_w and L_x. The lens increases the raw density by 86% and decreases the storage cost by 9%.

Figure 5.7 illustrates the dependence of the storage density of the module illustrated in Figure 5.3c (with and without a lens) upon the two independent parameters (the pixel size b/ϕ and the 1-D number of pixels N). The corresponding contour plots are shown in Figure 5.8. The lower limit of the modulator/detector dimension b was decreased to 1 μm for these computations (the corresponding minimal pixel pitch b/ϕ is 2.9 μm), in order to illustrate the existence of an optimum pixel pitch around $b/\phi \approx 5.5$ μm. Larger pixel pitches do not take full advantage of the information packing potential of the optical system, while smaller pitches suffer from excessive diffraction spread (see §5.2.2). Technological limitations impose a minimum pixel pitch that is slightly larger than the optimum (see the discussion on page 133). This is why the optimum pixel pitch is 7.9 μm ($b = 2.75$ μm), the minimum allowed value.

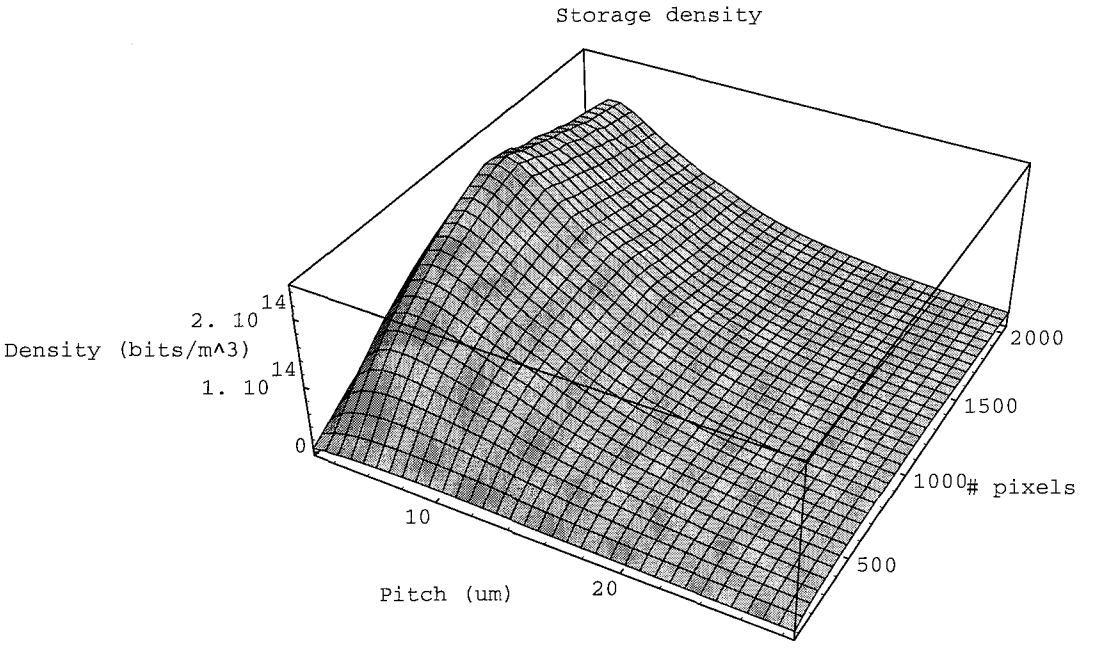
The storage cost (in dollars per raw megabyte) is plotted against the pixel size b/ϕ and the 1-D number of pixels N in Figure 5.9. We see that the storage cost is a steep function of the number of pixels. Large data pages ($> 512 \times 512$ pixels) are clearly desirable. The white sections in the regions of large b and large N in Figure 5.9 are the fobidden areas defined by

$$2.25 \left(\frac{Nb}{\phi} \right)^2 \geq A_{\text{OEIC,max}},$$

where the area of the OEIC would exceed technological limits (see the discussion on page 128). We argue that this limitation of the OEIC does not have a very negative effect on either the density or the cost of the modules. First, we notice that the maximum density, with or without a diffractive lens in the path of the signal beam,



(a)



(b)

Figure 5.7: Storage density of the module shown in Figure 5.3c, without (a) and with (b) a lens in the signal path.

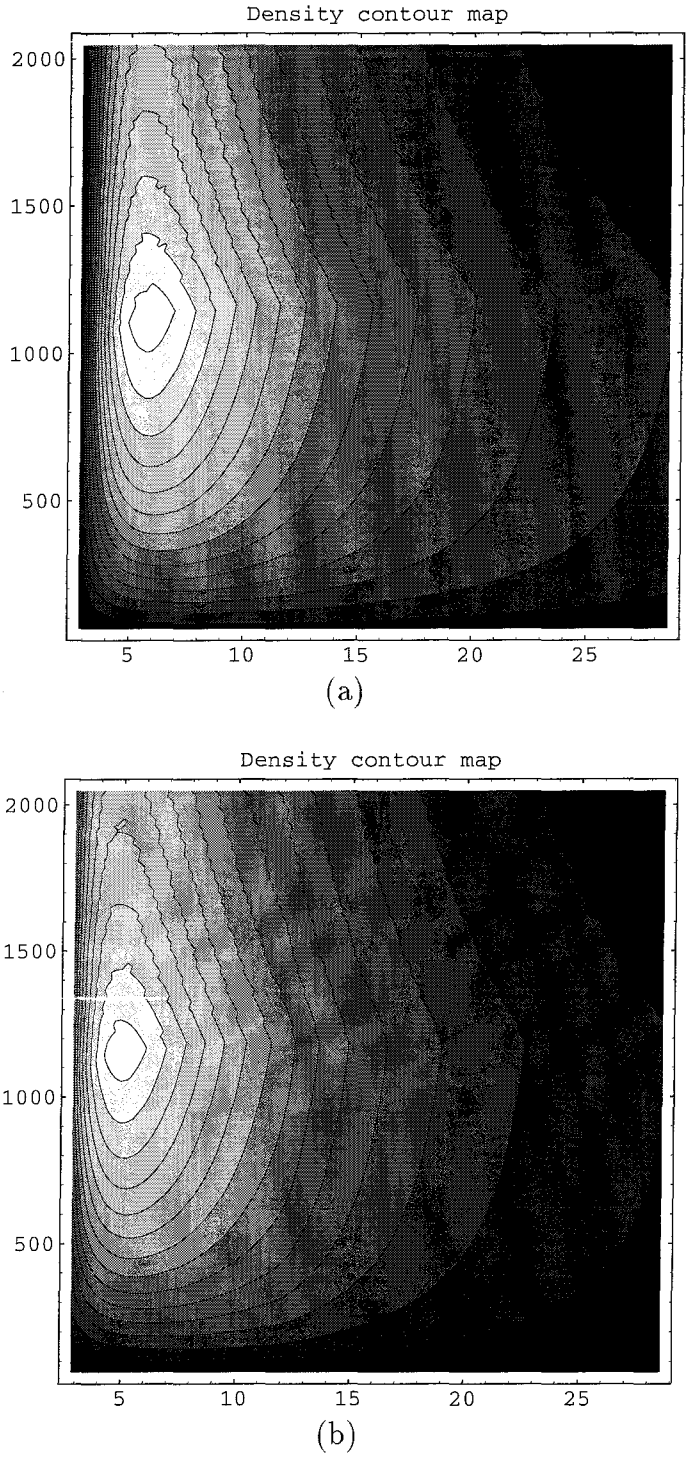
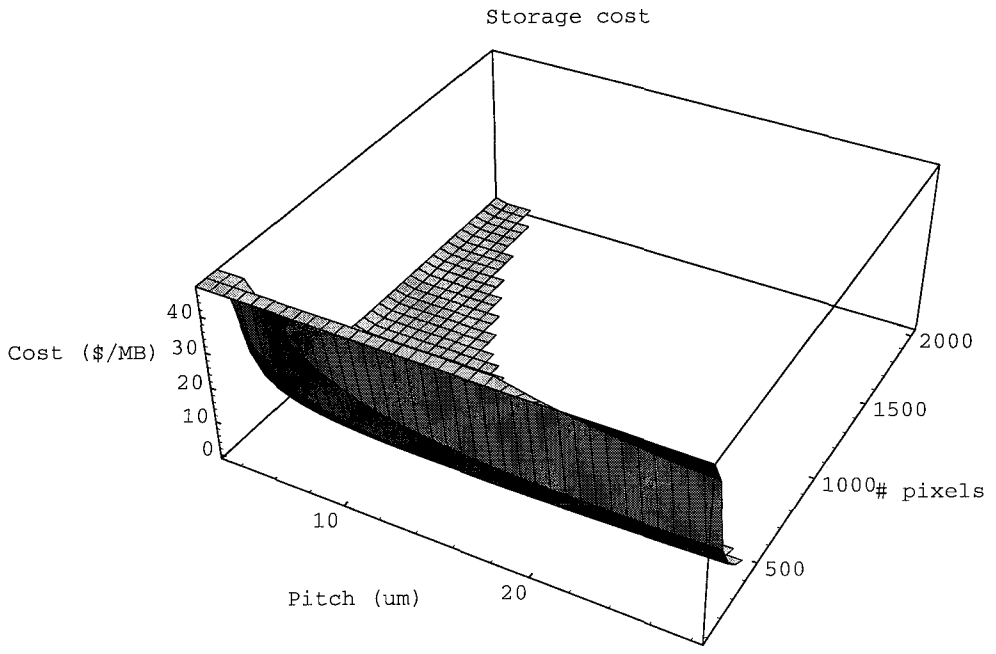
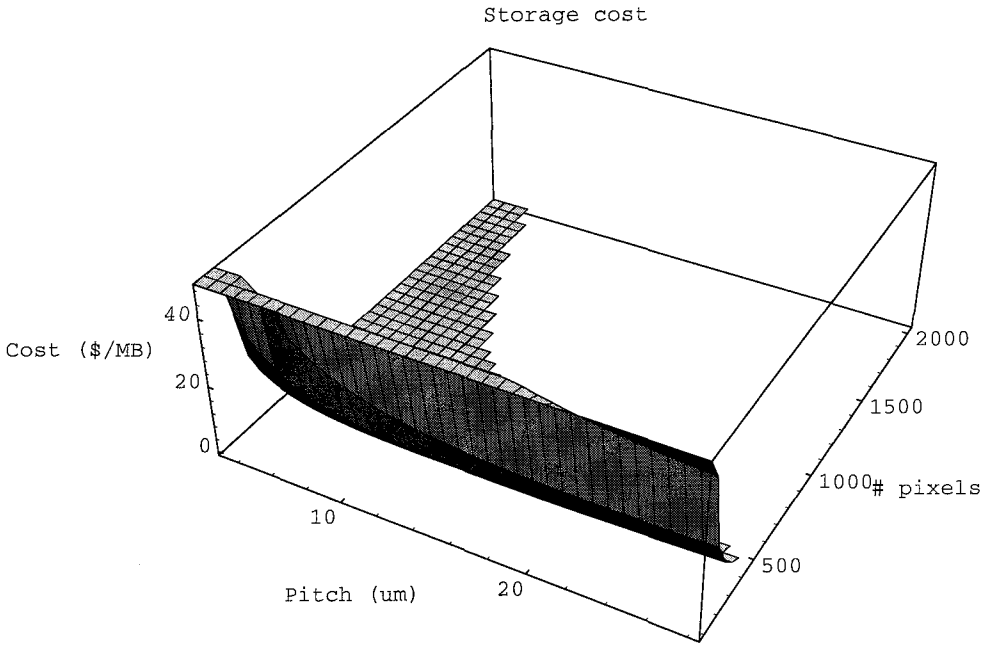


Figure 5.8: Storage density (contour plot) of the module shown in Figure 5.3c, without (a) and with (b) a lens in the signal path.



(a)



(b)

Figure 5.9: Storage cost of the module shown in Figure 5.3c, without (a) and with (b) a lens in the signal path.

is reached at a point within the "allowed" region, $2.25\,(Nb/\phi)^2 \leq A_{\mathrm{OEIC,max}}$ (see Figures 5.7 and 5.8). Note that the density optimization was carried out using an unconstrained OEIC die size. Second, in the case of the module without a lens, the minimum storage cost is also attained at a die size smaller than the limit (but very close to it); if we lifted the die size limitation, the optimum storage cost would not improve. The situation is different in the case of the module employing a lens, in which case the optimum storage cost is reached at the very boundary of the "allowed" region[18]. A bigger die could yield a somewhat lower storage cost, although the improvement would probably be minor, since the storage cost is a rather slow function of b and N around the boundary at large N.

Comparison between modules providing highest density and lowest cost

Table 5.7 shows an interesting result: the storage costs of the two modules providing the highest storage densities (Figures 5.3a and 5.3c, both with a lens) are off by a factor of almost four, although their densities differ only by about 5%. We might intuitively expect a stronger correlation between density and cost, since the cost of most elements is a monotonically increasing function of their area or volume (see §5.2.5).

Table 5.11 enumerates the properties of the two modules, and Table 5.12 shows their cost breakdowns. The tables show that these two modules represent two opposite, and almost equally successful, strategies to achieve a high storage density. The architecture shown in Figure 5.3a minimizes the module's volume by eliminating as many components as possible and by employing relatively small data pages. Power losses (due to the double pass of the readout reference through the recording medium) and diffraction efficiency losses (due to the reduced overlap between the holograms and the readout reference) are the price to pay for the compactness of the module. The result is that the small module also has a low capacity, because the number of pages M is limited by the weakness of the reconstructed data pages upon readout.

In contrast, the architecture illustrated in 5.3c eliminates the most severe power

[18]Within the granularity of the optimization grid.

Symbol	Description	Fig. 5.3a with lens	Fig. 5.3c with lens
	Cost per raw megabyte	\$2.607	\$0.687
b	Modulator/detector size[a]	2.75 μm	2.75 μm
	Pixel pitch	7.86 μm	7.86 μm
M	Number of holograms	608	1454
N	1-D number of pixels per page	504	1184
Θ	Angular range	5.9°	6.0°
	Raw module capacity (bits)	1.54×10^8	2.04×10^9
$F/\#$	F-number of lens	2.24	2.24
L	1-D dimension of pixel array	4.0 mm	9.3 mm
L_w	Dim. of BS3	4.5 mm	10.5 mm
L_x	Max. transverse dim. of signal	4.8 mm	11.3 mm
L_y	Dim. of reference beam splitters	4.0 mm	9.3 mm
L_z	Dim. of recording medium[b]	4.4 mm	10.4 mm
V	Volume of module	0.72 cm^3	10.07 cm^3
	Raw storage density[c]	2.14×10^{14}	2.03×10^{14}
A_{OEIC}	Silicon area of DHR OEIC	0.35 cm^2	1.95 cm^2
	Modules per laser	1	1

[a] See Figure 5.4.
[b] Along the propagation axis of the signal beam.
[c] In bits/m^3.

Table 5.11: Properties of the modules providing the highest densities (Figures 5.3a and 5.3c, both with a lens).

Description	Fig. 5.3a with lens	Fig. 5.3c with lens
Recording medium	$1.62	$12.65
DHR OEIC	$10.29	$34.21
Beam steerers[a]	$7.84	$86.54
Beam splitters[b]	$6.00	$10.07
Liquid crystal polarization rotators[c]	$2.25	$3.00
Source (laser & collimator)	$14.00	$14.00
Diffractive lens	$2.00	$2.00
Assembly	$4.50	$4.50
Total	**$48.00**	**$166.97**

[a]The modules illustrated in Figures 5.3a and 5.3c contain one and two beam steerers, respectively.

[b]The modules illustrated in Figures 5.3a and 5.3c contain three and five beam splitters, respectively.

[c]The modules illustrated in Figures 5.3a and 5.3c contain three and four polarization rotators, respectively.

Table 5.12: Cost breakdown of the modules providing the highest densities (Figures 5.3a and 5.3c, both with a lens).

limitations mentioned above by adding one beam steerer and two beam splitters. The additional readout power allows M and N to be increased. The density achieved by this module is almost as high as that obtained by the module of Figure 5.3a, because its increased volume is matched by a commensurately improved capacity.

While the storage densities of the two module architectures differ by only 5%, the storage cost of the architecture shown in 5.3c is 3.8 times less than its competitor's. Figure 5.3a's minimalist approach is expensive because the cost of the tiny module is dominated by fixed costs: optical power source, support components and assembly.

Effect of improved $M/\#$

The values of $(M/\#)^*$ assumed in our analysis are lower than the best numbers reported in the literature because we assumed the operating wavelength of the laser diode did not coincide with that at which the recording material is the most sensitive. Inexpensive and powerful laser diodes currently operate in the red or infrared; however, the peak sensitivity of well known, inexpensive recording media such as $LiNbO_3$ tends

to lie at shorter wavelengths, in the blue or green.

Nevertheless, ongoing developments may bridge this "wavelength gap." We can hope that recording materials and laser diodes will converge to a common optimal wavelength. As explained in §1.2.4, the optical disc market is pushing the development of laser diodes operating at shorter wavelength, which allow denser disc storage. There are also significant ongoing efforts to develop materials and recording techniques that allow holographic storage systems to work at longer wavelengths (see §1.2.3). The main result of this convergence in wavelength would be an increased $(M/\#)^*$. We recall from earlier discussions that many important performance parameters exhibit a quadratic dependence upon $(M/\#)^*$. Tables 5.13 and 5.14 show the effects of a hypothetical 45% increase of $(M/\#)^*$ on the module of Figure 5.3c, with a lens in the signal path, optimized for cost. As a result of the increased $(M/\#)^*$, the storage cost decreases from \$0.635/MB to \$0.520/MB. While the module's capacity and cost effectiveness were limited by the available diffracted power at $(M/\#)^* = 3.4$ cm^{-1}, they become limited by the 12° angular range of the beam steerer at $(M/\#)^* = 5.0$ cm^{-1}. If this angular range is increased to 15° (i.e., $\Theta_{max} = 7.5°$, all other things being equal, the storage cost decreases to[19] \$0.446/MB at $(M/\#)^* = 5.0$ cm^{-1}.

5.2.7 Limitations of the density and cost models

We believe that the models decribed in §§5.2.1–5.2.5 capture most of the effects governing the storage density and cost achievable using integrated modular holographic memory. Nevertheless, some effects are currently unaccounted for.

Reflections The models currently do not account for reflections at the interfaces between module components. It is assumed that appropriate anti-reflection coatings are employed.

Diffraction efficiency of beam steerer While well designed liquid crystal beam steerers deflect most of the incident light in the desired direction, a small fraction

[19] Or \$0.497/MB without a lens in the path of the signal beam. The maximum raw storage density is 2.61×10^{14} bits/m^3 with a lens, or 1.38×10^{14} bits/m^3 without a lens.

Symbol	Description	$(M/\#)^* = 3.4$ cm^{-1}	$(M/\#)^* = 5.0$ cm^{-1}
	Cost per raw megabyte	\$0.635	\$0.520
b	Modulator/detector size[a]	2.75 μm	2.75 μm
	Pixel pitch	7.86 μm	7.86 μm
M	Number of holograms	1468	1797
N	1-D number of pixels per page	1464	1464
Θ	Angular range	4.9°	6.0°
	Raw module capacity (bits)	3.15×10^9	3.85×10^9
$F/\#$	F-number of lens	2.24	2.24
L	1-D dimension of pixel array	11.5 mm	11.5 mm
L_w	Dim. of BS3	13.0 mm	13.0 mm
L_x	Max. transverse dim. of signal	13.9 mm	14.0 mm
L_y	Dim. of reference beam splitters	11.5 mm	11.5 mm
L_z	Dim. of recording medium[b]	12.6 mm	12.8 mm
V	Volume of module	17.7 cm^3	17.86 cm^3
	Raw storage density[c]	1.78×10^{14}	2.16×10^{14}
A_{OEIC}	Silicon area of DHR OEIC	2.98 cm^2	2.98 cm^2
	Modules per laser	1	1

[a] See Figure 5.4.
[b] Along the propagation axis of the signal beam.
[c] In bits/m^3.

Table 5.13: Properties of the module of Figure 5.3c, with a lens and optimized for cost, at two values of $(M/\#)^*$.

Description	$(M/\#)^* = 3.4$ cm^{-1}	$(M/\#)^* = 5.0$ cm^{-1}
Recording medium	\$21.33	\$21.73
DHR OEIC	\$49.66	\$49.66
Beam steerers[a]	\$132.32	\$132.32
Beam splitters[b]	\$11.50	\$11.50
Liquid crystal polarization rotators[c]	\$3.00	\$3.00
Source (laser & collimator)	\$14.00	\$14.00
Diffractive lens	\$2.00	\$2.00
Assembly	\$4.50	\$4.50
Total	**\$238.30**	**\$238.71**

[a]The modules illustrated in Figures 5.3a and 5.3c contain one and two beam steerers, respectively.

[b]The modules illustrated in Figures 5.3a and 5.3c contain three and five beam splitters, respectively.

[c]The modules illustrated in Figures 5.3a and 5.3c contain three and four polarization rotators, respectively.

Table 5.14: Cost breakdown of the module of Figure 5.3c, with a lens and optimized for cost, at two values of $(M/\#)^*$.

of the input light power is steered in undesired directions[69]. The effect is most noticeable at large steering angles; it is one of the reasons why the angular range was limited to $\Theta_{\max} = 6.0°$.

Error control and formatting overhead In a practical system, a fraction of a holographic module's storage capacity would likely be reserved for error control and formatting purposes. While the formatting overhead would probably be minimal, the optimum rate of error control codes in this application is still an open question (see §5.3.2).

Relationship between overlap and selectivity As discussed in §5.2.3 (see Figure 5.5), the shift between the reference beams used for recording and readout reduces the amount of power diffracted into the conjugate signal beam upon readout. The reduced interaction volume between the conjugate reference and the holographic grating also weakens the angular selectivity of the holograms. This latter effect is currently unaccounted for. The angular spacing of holograms

dictated by equations 5.25 and 5.26 may give rise to excessive cross talk. Therefore, a larger angular separation may be necessary, in which case our models would overestimate the storage capacity of the holographic modules.

Expanding signal beam In our analysis of the overlap between the conjugate reference and the holographic grating (see again Figure 5.5), we do not account for the effects of the diffraction spread or focusing of the signal beam. The signal beam is assumed to have an approximately constant width L throughout the recording medium.

Position of beam steerers The storage density of the module architecture using two reflective beam steerers (Figure 5.3d) could be slightly improved if the beam steerers were affixed to the top faces of BS2 and BS5[20], instead of their external lateral faces.

Fringes In the modules employing a single beam steerer (Figures 5.3a and 5.3b), the counter-propagating reference beam necessary to obtain the desired conjugate reconstruction is obtained by letting the main reference beam bounce off the mirror surface at the bottom of the recording material. The interference of the counter-propagating beams will give rise to a fringe pattern, whose effects are currently not modeled.

Lens and non-co-located modulators and detectors It is usually easier to fabricate spatially separated, rather than co-located, modulators and detectors on an OEIC. While conjugate readout would normally direct the reconstructed pixels toward the modulators instead of the detectors during readout operations, we have devised straightforward ways to ensure that the optically encoded data bits fall on the detectors, as desired, in any of the module architectures shown in Figure 5.3 (more on this in §6.2.4).

[20]The beam splitting axes would of course have to be rotated by 90°.

5.3 Comparison with other storage technologies; outlook

5.3.1 Performance comparison

The density and cost analyses presented in §5.2 have considerably evolved and undergone several iterations over a number of months. In comparison, the performance models are still rather embryonic. We are presently in the process of fleshing them out, with our colleagues at Caltech and Holoplex, as part of a study commissioned by DARPA. Nevertheless, we present here some preliminary estimates.

Intrapage access time

As long as the desired data is located in the last data page read from the holographic medium, it resides in the DHR's on-chip electronic buffer, which can be thought of as a cache for the holographic medium. In Chapter 3, we saw that both DHR-4 and the latest DHR design contain electronic buffers. In the case of DHR-4, the buffer is the static memory bit contained within each pixel; whenever the device is not detecting a data page, the optical detection train is disconnected from the memory bit, and the whole chip behaves as a (low-density) static RAM. In the case of the latest DHR design (using dynamic pixels), a separate DRAM array buffers the most recently accessed holographic page. Therefore, the intrapage access time is expected to be similar to the access time of a static RAM (DHR-4) or a dynamic RAM (latest DHR design) employing a similar process. The DHR device will, however, be somewhat slower because of added parasitic capacitances due to light shields and, in DHR-4, other circuits connected to the memory bit (liquid crystal driver, transmission gates connected to the optical detection train). We expect the intrapage access time of the latest DHR design to be on the order of a few tens of nanoseconds.

Interpage access time

If the desired data is not already cached in the DHR's on-chip electronic memory, a new page must be accessed. This requires the following steps:

1. Optically addressing the desired page;
2. Sensing the optical data page read out of the holographic medium;
3. Signal processing (e.g., correlated double sampling, thresholding);
4. Data transfer to the electronic buffer.

Since steps 1 and 2 must be temporally disjoint, their durations must be added. The duration of step 1 is almost equal to the switching time of the liquid crystal beam steerer, which we estimate at 1 ms. The duration of step 2 is the integration time required to obtain a sufficient signal-to-noise ratio. We argued in Chapter 3 that a 1 ms integration time is consistent with DHR circuitry and with the light levels in high-density, low-cost integrated modular holographic memory. Step 3 adds a trivial duration to the interpage access time, while step 4 can be performed while I/O pins are being driven. Therefore, the interpage access time is estimated at 2 ms. It is significantly lower than the access time of magnetic disks.

Data transfer rate during readout

The peak data transfer rate is the maximum instantaneous data transfer rate. In integrated modular holographic memory, a sensible definition is the inverse of the electronic bus cycle time. This transfer rate can be sustained as long as the data being read out is located in the on-chip buffer. Assuming the DHR's performance is comparable to that of current standard 50 ns DRAM ICs and that its data bus is 32-bit wide, its cycle time would be approximately 90 ns[21][114, for example] and its peak data transfer rate would be 356 Mb/s. Furthermore, nothing prevents the

[21] When a DRAM IC is labeled as a x ns device, we mean that its access time t_{RAC} (i.e., the time between the falling edge of the row address strobe $\overline{RAS}$ and the appearance of valid, stable data on the output pins) is x ns. The cycle time t_{RC}, i.e., the minimum delay between successive $\overline{RAS}$ falling edges, is longer. $\overline{RAS}$ must be held high during the $\overline{RAS}$ precharge time, t_{RP}.

sense amplifiers, column latches and decoders and I/O buffers associated with the DRAM buffer to implement an extended-data-out (EDO) page mode. Assuming that the DHR's performance is comparable to that of current 50 ns EDO DRAM ICs (e.g., Micron Technology's MT4LC16M4G3/H9[114]) and that its data bus transfers 32 bits in parallel, its peak data transfer rate would increase[22] to 1.6 Gb/s. However, this peak rate is only obtained as long as data transfers are done to or from a single row (somewhere between 1024 and 2048 bits) of the DRAM buffer.

The sustained data transfer rate is the data transfer rate that can be achieved over an arbitrarily long number of cycles. The computation of a random-access sustained data transfer rate would require knowledge of the application-dependent frequency at which new pages need to be retrieved from the holographic medium. The necessary research and computations have not yet been performed; they will be part of the more complete study mentioned above. In the meantime, we compute a sustained sequential-access data transfer rate, which applies well to tasks such as multimedia retrieval. If we assume once again that the DHR's electronic performance is comparable to that of current 50 ns EDO DRAM ICs, that its data bus is 32-bit wide, and if we further take $N = 1464$ (in agreement with Table 5.9), the time it takes to electronically read an entire page out of the DHR's DRAM buffer is less that the optical page access time (address setup time plus integration time). The sustained sequential-access data transfer rate is thus optically (rather than electronically) limited. It is given by

$$\text{SSADTR} = \frac{N^2}{t_{\text{opt acc}}}, \tag{5.39}$$

where $t_{\text{opt acc}} = 2$ ms is the optical access time (see discussion above). We obtain SSADTR = 1.1 Gb/s, or 67% of the peak data transfer rate.

Recording performance

Recording a new data page will take several seconds, given the sensitivity of inexpensive recording materials in the red and the relatively low power output by the laser

[22]The cycle time is assumed to decrease to 20 ns[114].

Parameter	Value
Intra-page access time	30 ns–50 ns
Inter-page access time	2 ms
Peak data transfer rate	1.6 Gb/s
Sustained sequential data transfer rate	1.1 Gb/s
Recording rate	$\frac{2.1\times10^6}{t_{\text{rec page}}}$ 214 kb/s @ $t_{\text{rec page}} = 10$ s

Table 5.15: Estimated performance of a single integrated holographic memory module.

diode. Given the assumptions above, the electronic contributions to the recording rate will be negligible, and the recording rate is simply estimated as follows:

$$\text{RR} = \frac{N^2}{t_{\text{rec page}}}, \tag{5.40}$$

where $t_{\text{rec page}}$ is the page recording time. For example, if, as before, $N = 1464$, and if $t_{\text{rec page}} = 10$ s, we obtain RR = 214 kb/s.

Summary of module performance

Table 5.15 summarizes the rough performance estimates obtained above.

Aggregate performance

We can envision holographic memory modules connected to an electronic backplane, as shown in Figure 5.2. As explained in §5.1.2, the same basic module architecture can be employed in applications requiring very different performance levels. The performance depends on the degree of parallelism supported by the electronic backplane. A relatively low-end board may only be able to access a single module at a time, whereas a high-end board may serialize and time- or wavelength- multiplex the outputs of several modules on a fiber or fiber bundle. This may require a hierarchichal arrangement of fast GaAs serializers and laser diode drivers.

We expect the data transfer rates (peak and sustained) attained by such a parallel board to be roughly the product of the number of modules and of the data transfer

rate of a single module, assuming the interconnections and the serialization process do not cause serious bottlenecks. For example, given the same hypotheses as above, a 16×16 board could reach a peak data transfer rate of 410 Gb/s[23] and a sustained sequential data transfer rate of 282 Gb/s. However, the access times would not be helped by the added parallelism. In comparison, the sustained data transfer rates of current high-performance magnetic disk systems are on the order of 2 Gb/s.

Summary

In conclusion, the projected performance of integrated modular holographic memory systems is intermediate between the performance of conventional primary (RAM) and secondary (disks) storage. Integrated modular holographic memory offers read transfer rates close to those of DRAM, but access times closer to those of magnetic storage. However, pending improvements in material sensitivity, the recording performance of holographic systems is expected to lag behind both DRAM and magnetic storage.

5.3.2 Competitiveness

The optimization results presented above only take on their full meaning when they are compared to the density and cost of competing storage technologies, mainly dynamic RAM (DRAM) and magnetic storage. The raw storage density of integrated modular holographic memory is about twice as high as the net storage density of conventionally packaged, mainstream DRAM. For example, the net density of a typical 64 Mb DRAM IC in a 32-pin plastic SOJ package[114] is[24] 8.2×10^{13} bits/m^3. However, we must keep an eye on semiconductor memory technologies currently in development or in low-volume production, such as Irvine Sensors Corporation's DRAM Short Stacks (3-D DRAM multi-chip modules offering net densities as high as 2.6×10^{14} bits/m^3). Magnetic hard disk drives[25] featuring a net density of 4.4×10^{14} bits/m^3 have also begun shipping[115]. In comparison, the densities in Table 5.7 are *raw* density pro-

[23]Such a bandwidth would likely require 20 fibers or more.

[24]The net capacity of the IC is 6.71×10^7 bits, and the dimensions of its package are approximately 21.0 mm $\times$ 11.2 mm $\times$ 3.5 mm.

[25]The Toshiba HDD-2716 2.16 GB drive.

jections that need to be scaled down to account for error control codes and other control information. Note also that our density definition does not take the volume of the optical power source (laser and collimator) into account, primarily to keep laser sharing as an option. Laser sharing reduces cost and increases density, at the cost of a reduced board-level parallelism (performance). In conclusion, it seems that integrated modular holographic memory is not likely to possess a strong advantage over competing technologies on the front of storage density.

Fortunately, assuming the validity of our cost models, it appears that integrated modular holographic memory systems can be cost-competitive. In general, the performance of integrated modular holographic memories is expected to be intermediate between that of DRAM and that of magnetic storage. For example, the sustained read data transfer rate of integrated modular holographic memory systems rivals that of DRAM, and their access time is competitive with that of magnetic disks (but several orders of magnitude higher than DRAM access times). Since integrated modular holographic memory systems are not expected to provide a higher performance than DRAM, we must find a module architecture that leads to a cost per megabyte lower than that of DRAM. The higher cost of integrated modular holographic memory compared to magnetic storage must also be justified by a commensurate performance advantage.

As of this writing[26], the most economical type of DRAM is 16 Mb[27] ICs. The denser 64 Mb chips are still in the early stages of their product life cycle and are more expensive. The high-volume price of 16 Mb DRAM ICs underwent a sharp decline at the end of 1996, and appears to have stabilized at approximately \$5.80. The corresponding storage cost is \$2.90/MB = 0.035 millicent/bit, or approximately 3.2 times the cost forecasted for 1995 in the National Technology Roadmap for Semiconductors[111]. The NTRS forecast is shown using circles (o) and continuous lines in Figure 5.10. The storage cost is expected to drop by approximately 60% every three years:

[26]January 21, 1997.

[27]Following industry conventions, Mb stands for megabit and MB stands for megabyte (8 megabits).

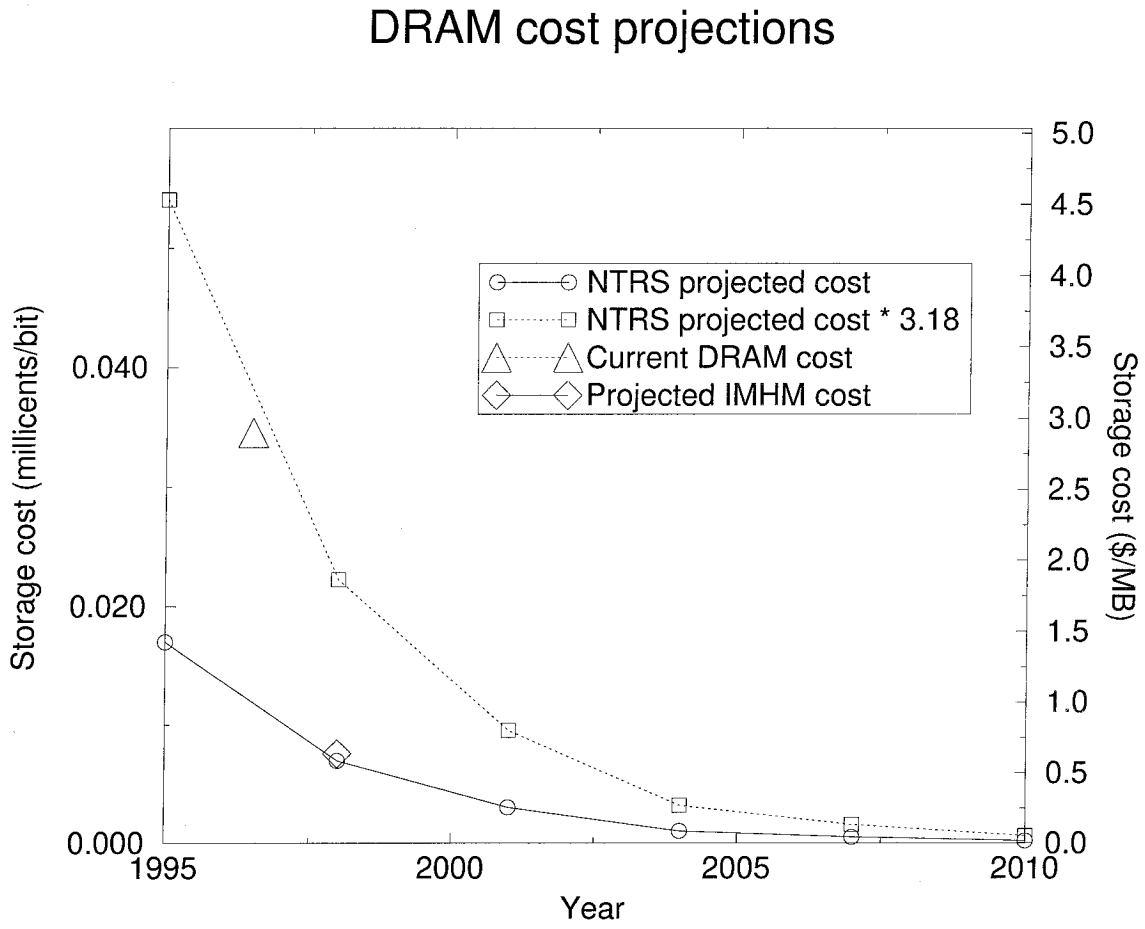


Figure 5.10: DRAM cost projections.

$$s(y) = s(1995)\, r^{(y-1995)/3}, \tag{5.41}$$

where $s(y)$ is the storage cost in year y and $r \approx 0.41$. The DRAM storage cost predicted by equation 5.41 for January 1997[28] is 0.011 millicent/bit = \$0.913/MB. The actual cost is indicated by a triangle ($\triangle$) in Figure 5.10. The DRAM market is known for its volatility. However, if we assume that the market continues to allow DRAM prices 3.2 times higher than the SIA's 1994 projections, the DRAM storage cost would follow the upper curve in Figure 5.10 (rectangles ($\square$) and dotted lines). Our cost projection for integrated modular holographic memory (IMHM) in represented by a diamond ($\diamond$) in Figure 5.10. Note that this storage cost does not take into account the overheads associated with error control codes and formatting information. Optimum error control schemes for holographic systems are still in active development, notably by Dr. Mark Neifeld's group at the University of Arizona[116, 117]. We see that the projected cost is substantially less than the upper DRAM cost curve; it is almost as low as the NTRS's initial projected cost for DRAM in 1998. Furthermore, as discussed under *Effect of improved M/#* in §5.2.6, developments in the fields of recording material and visible laser diodes could significantly bolster integrated modular holographic memory's cost competitiveness. Moreover, the cost of the modules could be decreased by sharing a laser between several modules[29]. This allows a compromise between cost and performance (parallelism).

The storage cost of magnetic disks currently varies between \$0.10/MB = 0.00119 millicent/bit and \$1.50/MB = 0.0179 millicent/bit, depending on the performance and format of the drive[115]. Compact and rugged drives for laptop computers are more expensive, as are top-performance models. Given the considerable performance lead of integrated modular holographic memory systems, we feel that the latter technology deserves its place in the sun in the storage cost gap between DRAM and magnetic storage.

[28] We assume that the predictions for a given year are effective in the middle of the year, i.e., around July 1.

[29] However, the laser accounts for less than 6% of the module cost (see Table 5.10).

5.3.3 Outlook

We expect the capacity and performance of integrated modular holographic memory systems to improve over the coming years because of developments in the fields of laser diodes, recording materials[30] and non-volatile rewritable volume holography. Laser diodes operating at shorter wavelengths will make it possible to reap the performance potential of well characterized, inexpensive photorefractive materials such as $LiNbO_3$. The development of such laser diodes is driven by mass-market applications such as optical disk storage. Better materials will provide higher sensitivity, higher diffraction efficiency and longer dark decay, resulting in higher capacity and density and shorter integration time (access time). Non-volatile storage techniques such as gated recording and the use of different wavelengths for recording and readout (with suitable page formatting) will enable non-volatile readout, and thus substantially increase the refresh period in typical applications.

We also expect integrated modular holographic memory systems to benefit from advances in IC processing and circuit design. We recall from Chapter 3 the similarities between dynamic DHR pixels and DRAM cells; the DHR device can take advantage of compact 3-D capacitor structures and advanced sense amplifier designs, for example, to deliver higher data transfer rates and pixel densities. Improved smart pixel technologies are likely to increase the modulator fill factor. The same device may be used alternately as optical input and optical output, as is currently possible with multiple-quantum-well modulators in the infrared. In fact, such modulators are readily applicable in laboratory prototypes that employ gated recording.

Nevertheless, the extent of the potential leverage of semiconductor technology developments by integrated modular holographic memory technology is limited. While DRAM costs have not decreased as much as predicted by the National Technology Roadmap for Semiconductors[111], perhaps more because of market conditions than technical difficulties, technological advances closely follow the schedule set forth in the NTRS. Minimum feature sizes have shrunk below 0.35 μm, 64 Mb DRAMs are

[30] On these first two topics, see also §1.2.4, §1.2.3 and the section titled *Effect of improved M/#* in §5.2.6.

Technology	Cost	Cost percentage
Silicon VLSI	$44.66	18.7%
Optoelectronic devices	$152.32	63.9%
Optical components	$36.83	15.5%
Other	$4.50	1.9%
Total	**$238.20**	**100%**

Table 5.16: Cost breakdown of module with lowest storage cost, by technology.

in production, 256 Mb chips will soon be available, and the first 1 Gb DRAM IC has been demonstrated[118], uses a 0.18 μm CMOS process and is slated to be commercially available at the turn of the century, exactly as projected in the NTRS. The cell size of 256 Mb DRAMs is approximately 0.6 μm^2[119], down from 1.8 μm^2 in 64 Mb devices[120, for example]. Unfortunately, integrated modular holographic memory systems may not be able to fully leverage the rapidly shrinking feature sizes and expanding die sizes resulting from massive investments in silicon fabrication technologies. For example, we saw in the earlier parts of this Chapter that optical effects limit the maximum achievable storage density. Given our assumptions, the semiconductor technology that will be available in 1998 will come close to reaching the optimum density (see Tables 5.7 and 5.8 and Figure 5.8). The density could only be improved by 22% through a reduction of the pixel size. Furthermore, we saw in §5.2.6 that neither the storage cost nor the density of integrated modular holographic memory could greatly benefit from the projected increased semiconductor die sizes[31]. Note also that semiconductor processing (i.e., the silicon die of the OEIC) accounts for only a small fraction of the total optimized holographic module cost (see Tables 5.10 and 5.16).

[31]This is true under the assumptions of our analysis. Of course, improvements such as an increased fill factor or the availability of a more powerful laser could certainly increase the storage density.

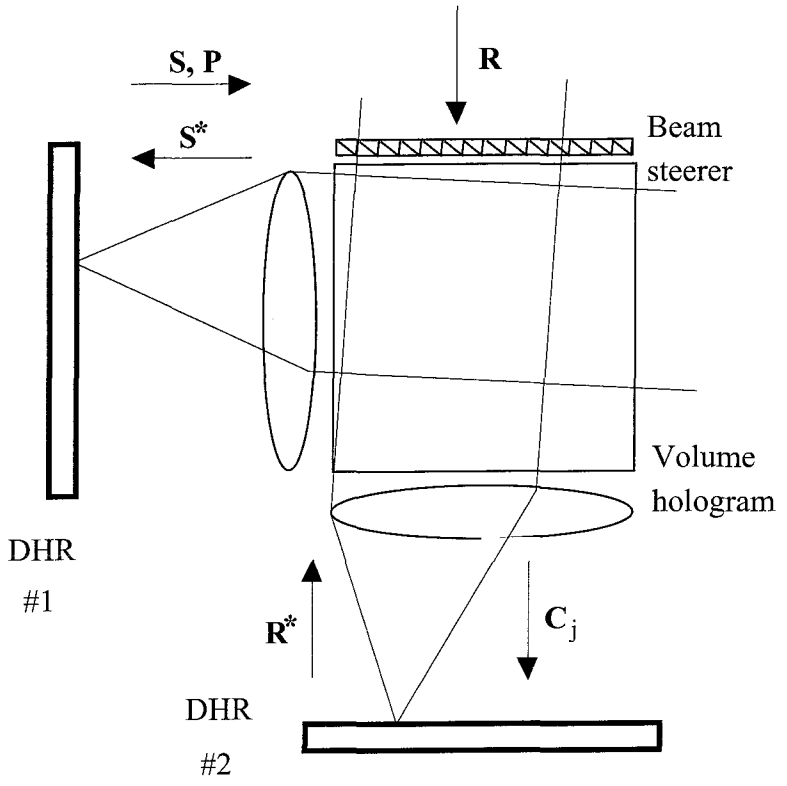


Figure 5.11: Holographic associative memory and correlation module.

5.4 Compact module for correlation and associative memory

We could not write a Chapter on integrated holographic module architectures without mentioning an elegant holographic information storage and processing module devised by our colleague George Barbastathis (see Figure 5.11). This multi-function integrated module works as

- a regular holographic memory;
- a correlator;
- an associative memory or content-addressable memory.

We describe the structure of the module, and then explain its operation in each mode.

5.4.1 Structure of the module

The module (see Figure 5.11) is comprised of a holographic medium in which volume holograms are stored, a beam steerer employed to angularly multiplex holograms, two DHR OEICs similar to those described in Chapter 3, two lenses, and two beam splitters (located between the lenses and the OEICs, and omitted from the diagram for the sake of simplicity and clarity) used to illuminate the reflective OEICs. The DHR OEICs are located one focal length away from the lenses. The beam steerer can focus the reference beam R on any of the pixels of DHR #2.

5.4.2 Operation as a regular holographic memory module

Holograms are stored in the recording medium by the interference of the signal beam S, modulated by DHR #1, with the reference beam R. Angular multiplexing is effected by the beam steerer. Note that the holograms store the Fourier transform of the data pages; in contrast, the lens-less architecture discussed previously stores the Fresnel diffraction pattern of the pages. To each hologram corresponds a unique deflection angle and a unique pixel on DHR #2.

To retrieve a hologram, the corresponding pixel on DHR #2 is turned on, while all other pixels are off. The location of the pixel addresses the desired data page. When the OEIC is illuminated[32], only this one pixel has a significant reflectance. The beam emanating from it is collimated by the lens above DHR #2, and becomes the conjugate of the recording reference of the desired data page. The desired reconstruction self-focuses back on DHR #1, where it is sensed and processed.

5.4.3 Operation as a correlator

The same module architecture can also be operated as a correlator (see §1.1.4). The correlation templates are first stored in the recording medium, using the procedure outlined above. An input pattern P is displayed on DHR #1. The adjacent lens

[32]A direct illumination path (omitted from Figure 5.11 for the sake of simplicity) is provided for each DHR.

Fourier-transforms the image. The interaction of this signal beam with the volume hologram computes the desired overlap integrals, one for each stored template. The correlation peaks are focused on DHR #2 by the lens at the bottom of the recording medium.

5.4.4 Operation as an associative memory

The two operating modes described above (regular memory and correlator) can be combined to form an *associative memory* or a *content-addressable memory.* The input of such a system is a pattern to be matched to the templates stored in the memory, rather than a simple address, such as an angular setting for the beam steerer. The goal of the memory is to output the stored template that most closely matches the input, according to a predetermined criterion. In our case, this criterion is the correlation integral or inner product. Such a system can restore a distorted input image that is part of a database, or retrieve a complete template based on a partial image, for example. It can also store associations between patterns. Consider the simple case where binary associations must be performed, for example between Arabic and Chinese numerals. Each stored template is divided into two tiles, one displaying an Arabic numeral, and the other, its Chinese equivalent. Given only a single Arabic numeral, the associative memory output the corresponding pair of numerical representations. The memory is robust to noise and distortions in the input image, because the correlation function is itself resilient to noise and distortions.

The associative or content-addressed recall proceeds in two steps. First, an input pattern P is displayed on DHR #1. The inner products between P and each of the stored template are computed as outlined above (the inner product is the value of the correlation integral at the origin (no shift)), and sensed by DHR #2. The modulator of the pixel that detected the largest inner product in turned on, while all other modulators are turned off. The operation requires the addition of a *winner-take-all* function to the DHR OEIC. The winner-take-all computation can be performed in analog VLSI with complexity proportional to the number of inputs[121]; an optoelectronic winner-

take-all integrated circuit using LCOS modulators has been demonstrated[122].

The second step of the process is identical to conjugate readout in the regular memory mode. DHR #2 is illuminated; a diverging wave emanates from its single "on" pixel, is collimated, and produces a self-focusing conjugate signal beam that is detected by DHR #1.

5.4.5 Hologram refreshment

Finally, the module can periodically refresh its stored templates or data pages, by successively using the readout and recording parts of the regular memory mode, as described in §5.4.2.

The main challenge in practically implementing this module would likely be the high insertion loss of DHR #2 during conjugate readout and associative recall. During these operations, a reference beam is generated by the reflection of an illuminating beam on DHR #2, in which all modulators but one are in the "off" state (i.e., not reflective). Most of the input light power is therefore wasted, because it impinges on "off" modulators or on non-reflecting areas of the OEIC.

5.5 Integrated module architectures employing non-angular multiplexing

In the preceding sections of this Chapter, we discussed at great length integrated holographic storage module architectures employing angular multiplexing. The reason for emphasizing this particular multiplexing technique is that it is the one that is best known and characterized, and most likely to quickly lead to successful laboratory prototypes.

Nevertheless, our modular and integrative approach is general. It applies to most known multiplexing methods. In what follows, we illustrate this point by showing and briefly describing module architectures employing wavelength, phase-code and shift multiplexing[102]. A detailed discussion of the issues involved in each architecture is

beyond the scope of this text. The reader will probably imagine a number of possible variations and challenges associated with each module.

5.5.1 Wavelength-multiplexed module

A wavelength-multiplexed[123] module architecture is shown in Figure 5.12. It employs the counter-propagating geometry that ensures an optimum wavelength selectivity. This remarkably simple module consists of a photorefractive crystal, a beam splitter and an optoelectronic integrated circuit. Holograms are written by illuminating the system from the left with the reference wave for writing (R_w). After propagating through the crystal and beam splitter, this beam generates a counter-propagating signal beam upon reflection and modulation by the OEIC, which displays the bit pattern to be stored. Different data pages are superimposed in the same volume and are recorded with slightly different wavelengths. During readout, the reference wave for readout, R_r, is shone alone; its wavelength is adjusted to match that employed to write the desired data page. After being reflected by the beam splitter, it propagates toward the left and becomes the conjugate of the reference beam employed to write the hologram. The resulting phase-conjugated reconstruction propagates back toward the OEIC, where it is detected and binarized.

A potential difficulty associated with the system shown in Figure 5.12a is reflections from the various interfaces (e.g., between the beam splitter and the storage crystal) during readout. These reflections could impede the detection of the weak hologram, unless very good anti-reflection coatings are employed. The system shown in Figure 5.12b avoids this problem by employing the 90° geometry. The main and conjugate reference beams are normally incident on the photorefractive crystal (PRC) from the left and the right, respectively. For example, they could be generated by two collimated laser diodes (LD1 and LD2) attached to the crystal. The signal beam comes from the left, bounces off the beam splitter, is modulated by the OEIC and reflected toward the crystal, where it interferes with the main reference beam (from LD1) to write holograms. Data pages are read out by illuminating the crystal with the

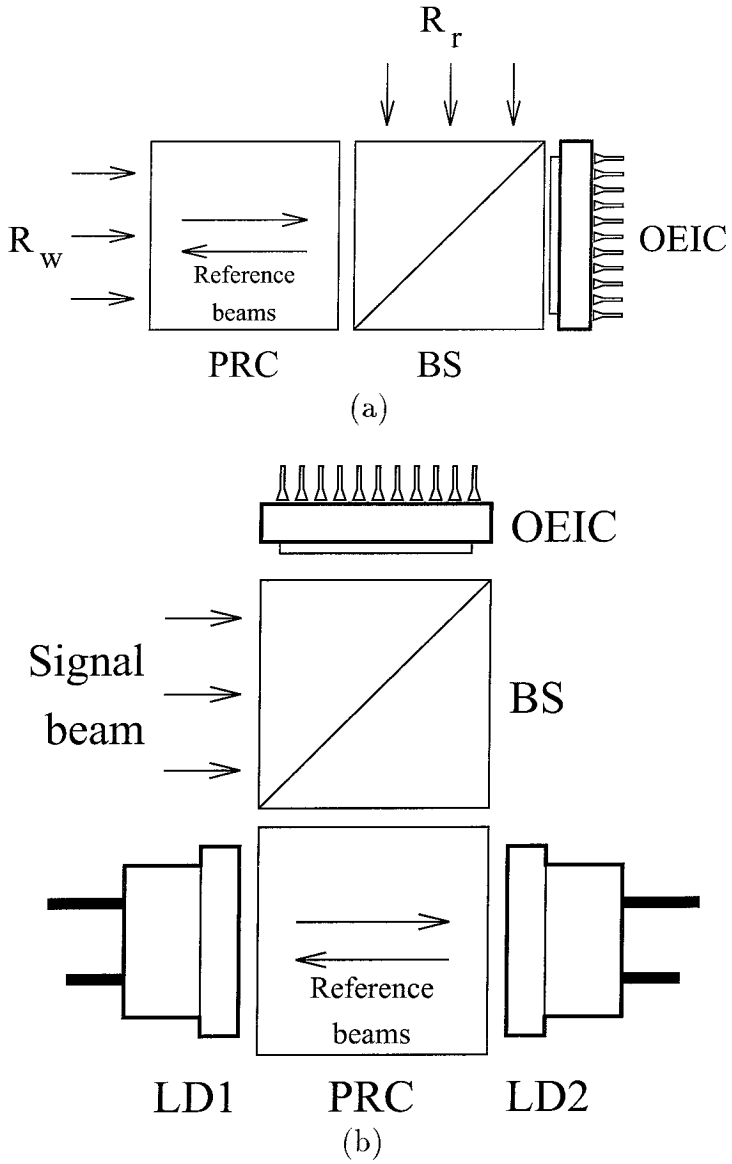


Figure 5.12: Wavelength-multiplexed storage module (a), comprised of a photorefractive crystal (PRC), a beam splitter (BS) and an OEIC. The reference beam for writing (R_w) illuminates the system from the left-hand side, and generates a counter-propagating signal beam upon reflection and modulation by the OEIC. A similar system employing the 90° geometry (b) avoids reflection problems, at the cost of a reduced wavelength selectivity.

conjugate reference (from LD2); the phase-conjugate reconstruction propagates back to the OEIC.

5.5.2 Phase-coded module

Figure 5.13a illustrates the second architecture, based on phase-code multiplexing[124, 125, 126]. The signal beam is generated in the same way as in Figure 5.12b. The reference beam illuminates a phase spatial light modulator (PSLM), which imparts a different phase delay to each pixel. This modulator is located in the front focal plane of lens L1. If the pixels are small enough, the light beam emanating from each of them can be thought of as a diverging spherical wave. The set of spherical waves becomes a collection of plane waves propagating a different angles upon traversing lens L1. All these plane waves are superimposed in the crystal. A different vector of phases, or phase code, is employed to record each data page. The set of phase codes is orthogonal. When retrieving data, the phase code employed to write the desired data page is displayed on the phase SLM. One could think this would produce a large number of superimposed holographic reconstructions, since the same set of plane waves was employed to record a large number of holograms. However, because of the orthogonality property of the phase codes, only the desired data page is reconstructed coherently; partial reconstructions of the undesired data pages associated with individual plane waves interfere destructively and cancel each other out. As in the module shown in Figure 1.7, the reconstructed data page would normally propagate away from the OEIC, toward the bottom of Figure 5.13a. However, each plane wave can be made to retrace itself by adding a second lens (L2) and a mirror (M) to form a 4-f system in the reference path (PSLM, L1, L2, M). The desired phase-conjugate reconstruction propagates toward the OEIC, where it is sensed.

While the system of Figure 5.13a possesses the main advantages associated with the other modules (combined data SLM/detector array, compact integration of OEIC, BS, PRC, L1 and L2), it is more difficult to compactly integrate the phase SLM (PSLM) and the mirror (M) with the rest of the module, since these elements must

be a focal length away from lenses L1 and L2, respectively.

A potential solution to this problem is shown in Figure 5.13b. All the elements of this system are in contact and attached to each other. The lenses have been eliminated. This storage module employs a slightly different kind of phase-code multiplexing. The pixel structure of the phase modulator divides the crystal into a number of slices. During readout, the modulator displays the phase vector employed to record the desired data page; the hologram "slices" corresponding to the desired data page add in phase, whereas those of other data pages add destructively due to the orthogonality of the phase codes. A self-retracing (phase-conjugate) reconstruction is obtained by conjugating the reference beam using a four-wave phase conjugator (a simple mirror would not work because the beams associated with each pixel of the phase SLM diffract). This architecture raises issues of its own, however. For example, two additional beams (the pump beams) must illuminate the module and be aligned with each other. Moreover, the power efficiency and the response time of the phase conjugator must be carefully optimized.

5.5.3 Shift-multiplexed module

Our last architecture is illustrated in Figure 5.14. This module employs shift multiplexing[127]. In this case, the reference wave is a converging beam obtained, for example, by inserting a lens (L1) in the path of the plane wave. Holograms are multiplexed by translating the recording medium perpendicularly to the propagation axis of the reference wave. In a typical system, the shift between holograms can be as small as a few micrometers. Self-retracing reconstructions are obtained by "folding back" the spherical wave using a mirror (M) positioned at its waist (focal point). This mirror could simply be a metallic coating on a face of the recording medium. The main drawback of this system is the relative motion it requires between the lens L1 and the rest of the module. It may be possible, however, to deflect the reference beam using a programmable diffractive optical element instead of resorting to mechanical motion.

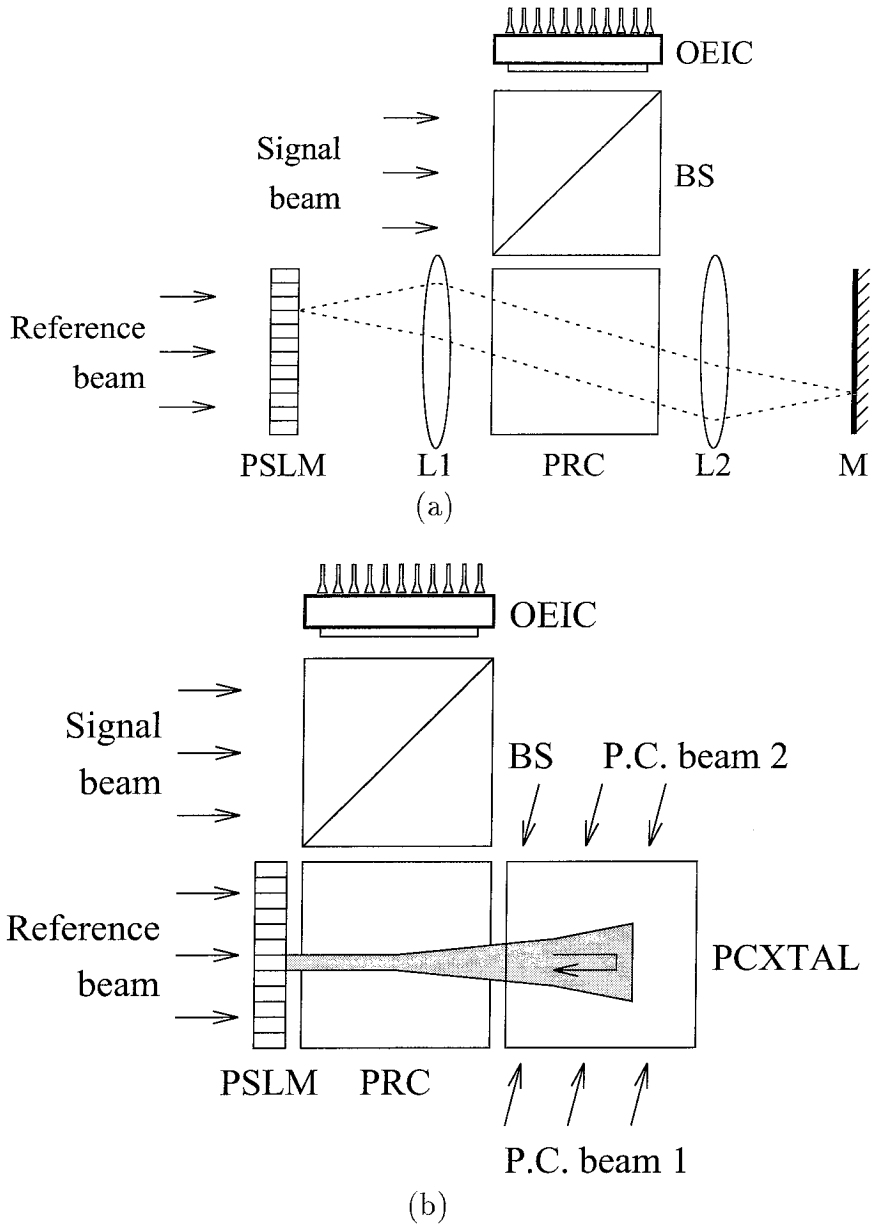


Figure 5.13: Phase-coded memory modules.

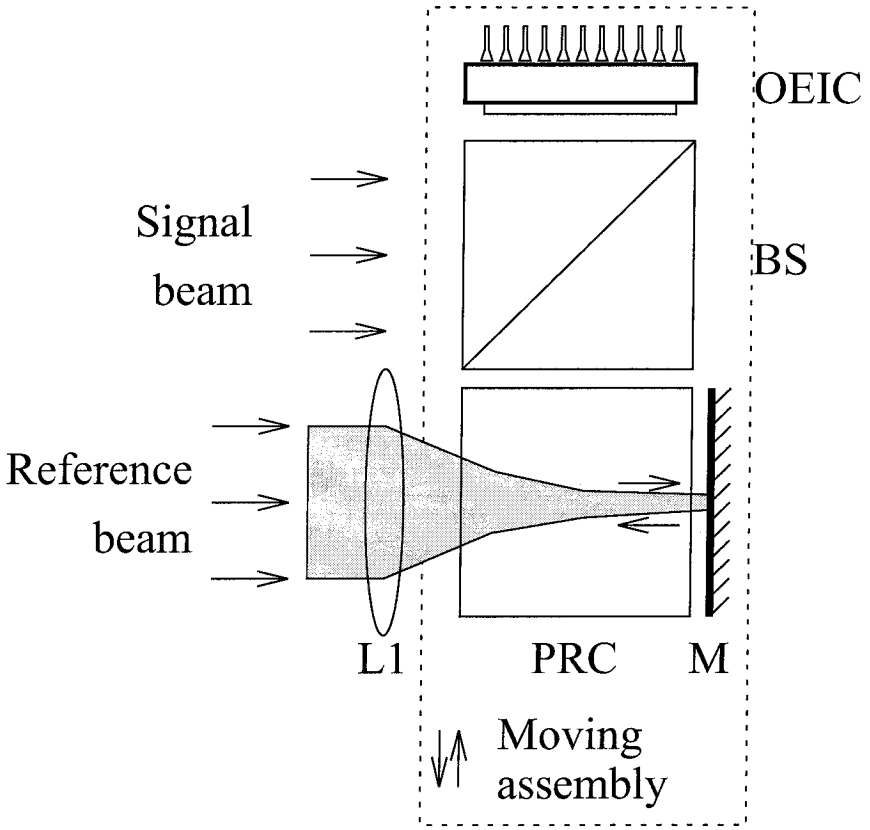


Figure 5.14: Shift-multiplexed module.

Chapter 6 Sustained holograms

6.1 Introduction

This Chapter describes experimental demonstrations of the main principles leading to the compact and modular read/write holographic memory discussed in §1.3 and in Chapter 5:

- conjugate readout;
- periodically refreshed holograms.

We are greatly indebted to our collaborators Ernest Chuang and George Barbastathis, who played essential roles in this work.

The experiments described in §6.2 were performed at the end of the summer of 1995, with our colleague George Barbastathis. They demonstrated conjugate readout using two different methods of obtaining the necessary mutually conjugate reference waves for recording and readout[76, 128]. They also demonstrated that the DHR-4 device can be employed to record high-quality photorefractive holograms.

More recently, in collaboration with Ernest Chuang, we demonstrated a laboratory prototype of a compact holographic storage module using a lens-less signal beam path. Holograms were recorded and retrieved by means of the DHR-4 OEIC. A hologram was sustained over 50 refresh/decay cycles[129].

6.2 Demonstration of conjugate readout

We saw in §1.3 that conjugate readout requires counter-propagating, mutually conjugate reference beams: one for recording, and one for readout. This requirement is independent of the multiplexing technique employed by the storage module; it applies not only to the angularly multiplexed module shown in Figure 1.7, but also to the

modules based on wavelength, shift and phase-code multiplexing presented in §5.5. All the modules considered in §5.2 (see Figure 5.3) employed a set of mirrors (beam splitters) to obtain the readout reference beam. A careful alignment of the mirrors, possibly complemented by corrections performed by a beam steering device, ensures that the readout reference is the phase conjugate of the recording reference. Alternatively, the readout reference can be produced by phase-conjugating the recording (forward) reference. Our early experiments demonstrated both approaches.

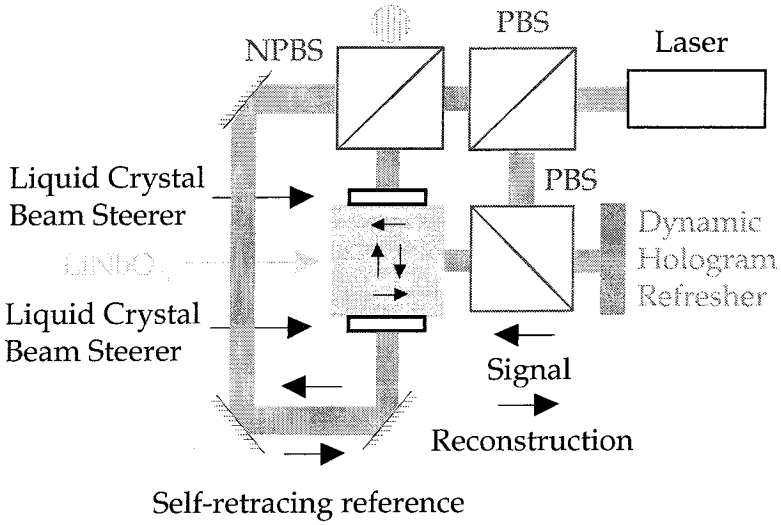
6.2.1 Mirror loop

Figure 6.1a shows a generic memory module architecture employing a set of mirrors and beam splitters to produce the two desired counter-propagating reference waves. Note how similar it is to the module illustrated in Figure 5.3c. The complete experimental setup employed to demonstrate this architecture is represented in Figure 6.1b.

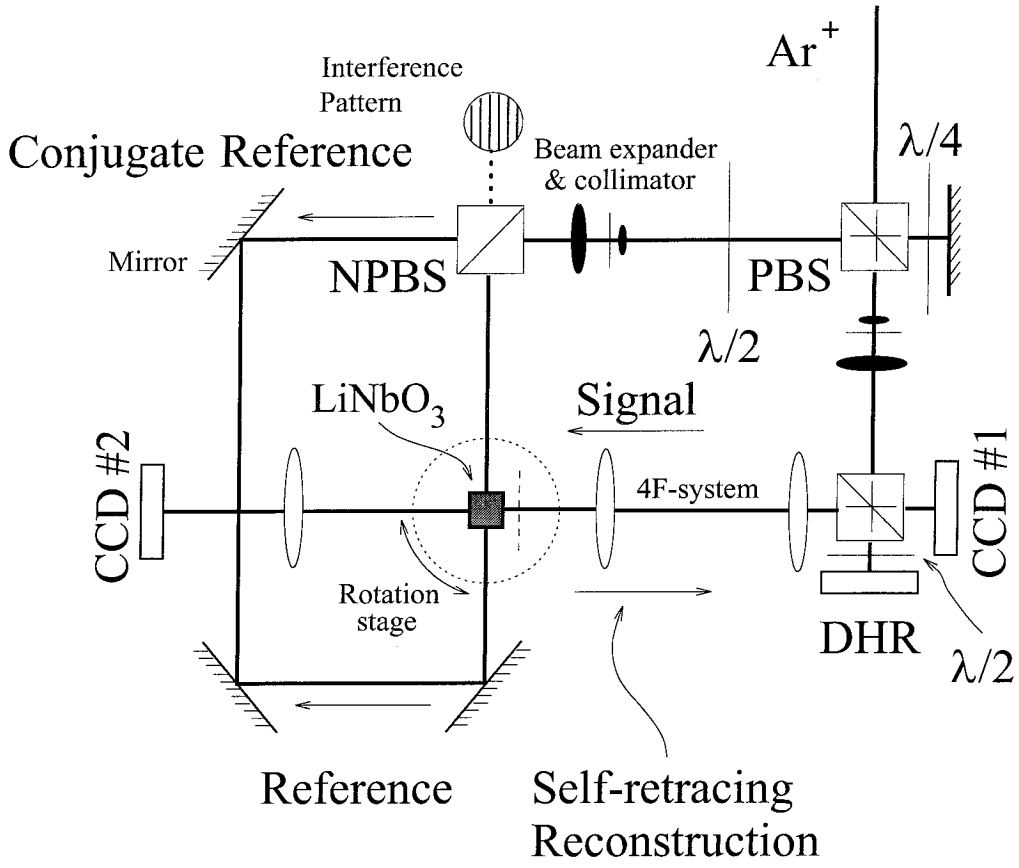
Generation of reference and signal beams

The input beam, produced by an Ar^+ laser ($\lambda = 488$ nm), entered the setup though a polarizing beam splitter (PBS) near the upper right corner of the diagram. The transmitted beam illuminated the DHR OEIC and became the signal beam. The reflected component[1] made a short round-trip to the right of the beam splitter before reaching the mirror loop to produce a pair of counter-propagating reference beams. The mirror on the right side of the PBS was mounted on a translation stage which allowed adjusting the length of the path traveled by the reference wave. The coherence length of the particular laser employed in this experiment was only on the order of a few centimeters. The signal and reference path lengths thus had to be carefully matched to ensure that a high-contrast interference pattern was obtained inside the recording medium ($LiNbO_3$:Fe). The quarter-wave plate, whose optic axis was oriented at 45° with respect to the plane of the figure, rotated the input polarization by 90° to ensure

[1] The wave reflected to the right is vertically polarized.



(a)



(b)

Figure 6.1: Generic conjugate readout setup using a mirror loop (a), and experimental setup (b).

that the beam reflected toward the left was transmitted by the PBS. The half-wave plate in the reference path rotated the polarization of the reference beam to a vertical state, as required for a strong interaction with the signal beam in the 90° geometry.

Modulation and imaging of the signal beam

Data was imprinted on the signal beam by a DHR-4 OEIC, near the bottom right corner of the schematic shown in Figure 6.1b. The modulated signal beam was reflected toward the recording medium by a beam splitter. A half-wave plate located between the beam splitter and the DHR rotates the polarization of the incident beam so that the beam impinging on the DHR is linearly polarized at exactly 45° with respect to both eigenaxes of the liquid crystal cell, in order to maximize contrast. The rotation is undone when the beam once again traverses the retardation plate after being reflected by the DHR. In order to allow for measurements and easy experimentation, the components were widely spaced in this early experimental setup. The DHR was imaged a few millimeters to the right of the recording medium (dashed line in Figure 6.1b) by a 4-F system. Holograms were therefore recorded in the Fresnel diffraction regime, as they would be in any of the module architectures illustrated in Figure 5.3.

The conjugate reconstructions were imaged on CCD #1 by the 4-F system. The forward reconstructions could also be captured using CCD #2.

Sagnac interferometer

The non-polarizing beam splitter and the three mirrors in the left half of either part of Figure 6.1 are responsible for generating the pair of conjugate reference beams required for conjugate readout. A very careful alignment is required to guarantee the Bragg matching of the readout (i.e., conjugate) reference beam with the diffraction gratings stored in the recording medium. The angular tolerance on the alignment of the conjugate reference in the plane of the figure is on the order of $\lambda/L \approx 0.003°$, where λ is the wavelength and L is the dimension of the interaction length[2]. Fortunately, the

[2]The angular tolerance in and out of the plane of the figure is considerably larger, on the order of $\sqrt{\lambda/L}$.

components can be aligned interferometrically. Note that the beam splitter and the three mirrors form a Sagnac interferometer[130, page 359]. The interference pattern appears in the back of the beam splitter, and can be employed to align the mirror loop before holograms are recorded.

6.2.2 Self-pumped phase conjugator

An alternative way of generating the desired counter-propagating and phase-conjugate reference beam required for conjugate readout is to employ a phase conjugator to cause the main reference beam to retrace its path. A generic four-wave mixing phase conjugator[25, Chapter 17] could be employed. However, a more compact and robust system can be obtained by using a self-pumped phase conjugator in the "cat" geometry[75]. The general architecture is illustrated in Figure 6.2a; the details of our experimental setup are illustrated in Figure 6.2b.

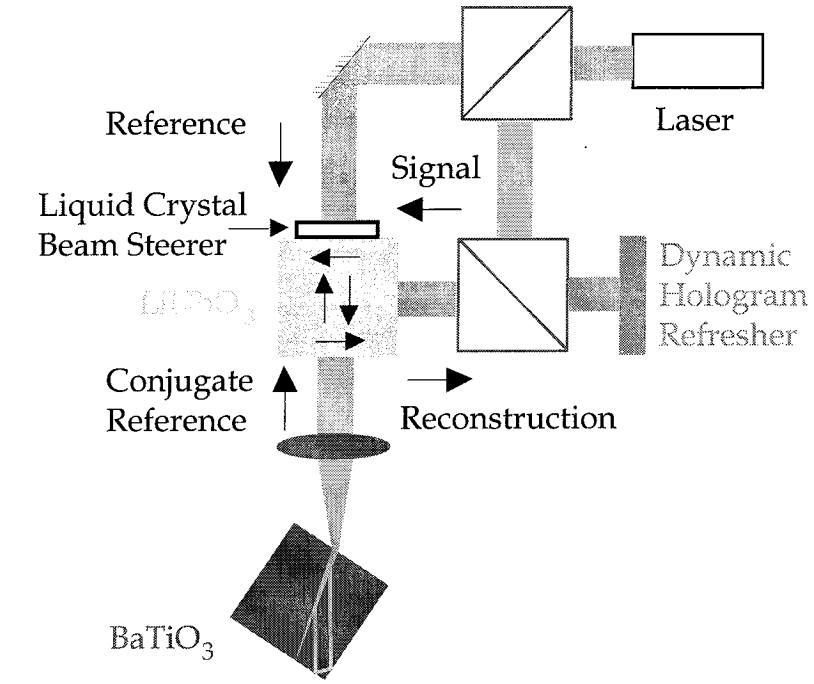
The right half of the experimental setup is identical to that of Figure 6.1b. The $BaTiO_3$ phase-conjugating crystal is located at the bottom of the reference arm. A lens focuses the incident reference beam to improve the efficiency of the phase conjugation process by increasing the local light intensity.

The system using a self-pumped phase conjugator (Figure 6.2) is arguably more elegant than its counterpart shown in Figure 6.1. Not only is it potentially more compact, but it is also considerably more robust to misalignments (e.g., components slightly misaligned during fabrication, effects of thermal expansion...). However, several issues would have to be addressed before this architecture becomes suitable to commercial production.

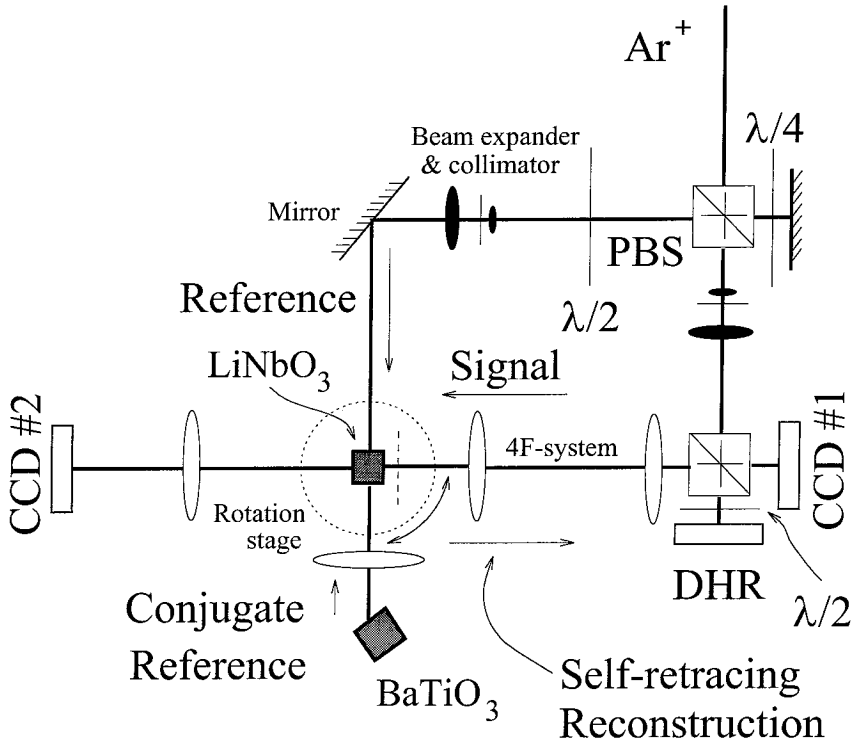
1. The reflectivity of the phase conjugator is considerably less than the efficiency of a mirror loop (see §6.2.3).

2. The phase conjugator requires a second crystal, which can be quite expensive.

3. The inter-page access time of the memory is increased by the settling time of the phase conjugation gratings.



(a)



(b)

Figure 6.2: Generic conjugate readout setup using a self-pumped phase conjugator (a), and experimental setup (b).

4. The efficiency and the response time of the phase conjugator depends on the input light intensity; at intensity levels compatible with laser diodes, the phase conjugator may be unreasonably slow or inefficient.

Issues 1 and 4 would be alleviated if a four-wave-mixing phase conjugator were substituted for its self-pumped counterpart in Figure 6.2. However, the compactness and robustness benefits of the phase conjugator approach would be eliminated, because of the need to route the two counter-propagating pump beams to the phase-conjugation crystal and to guarantee their alignment.

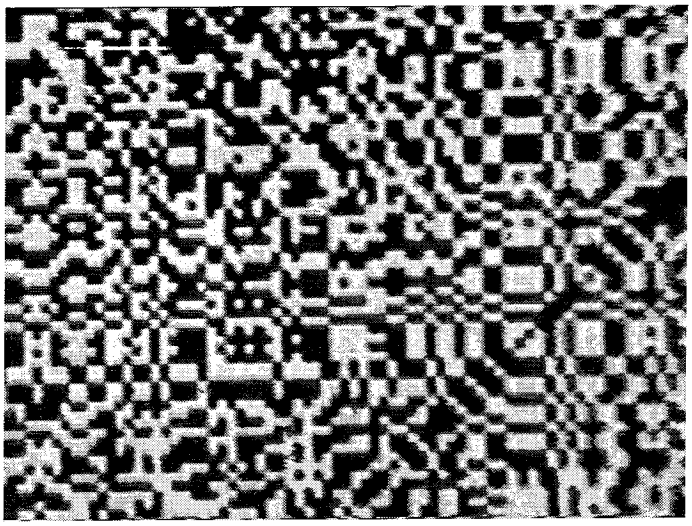
6.2.3 Experimental results

Both methods of generating a conjugate reference described in §§6.2.1 and 6.2.2 were experimentally demonstrated and employed to obtain conjugate holographic reconstructions[76]. Our initial experiments substituted simple black and white transparencies for the DHR OEIC; the setups were otherwise almost identical to those illustrated in Figures 6.1b and 6.2b. Figure 6.3 shows forward and conjugate reconstructions obtained in a setup using a self-pumped phase conjugator. Both reconstructions possess high quality. The conjugate reconstruction, however, exhibits a higher fidelity because phase conjugation undoes aberrations in the optical paths. The improvement is quantified in §6.3, which reports on more recent experiments. Similar results were also obtained with the mirror loop (Figure 6.1). The effective diffraction efficiencies[3] of the forward and conjugate reconstructions were respectively 4.0×10^{-4} and 4.5×10^{-5}. The reflectivity of the phase-conjugate mirror could be as high[4] as 30%, while its response time was less than 1 s.

The next step after these initial experiments was to demonstrate the recording of holograms of the DHR display. The experimental setup using a mirror loop, schematically shown in 6.1b, was employed for this purpose. Figure 6.4 shows the letters C, I and T displayed on a small portion of the DHR's pixel array, imaged through the $LiNbO_3$:Fe crystal and captured by CCD #2. High-fidelity holograms were success-

[3]Including losses in the imaging paths.

[4]At 7.8 mW input power, it dropped to 8%.



(a)

(b)

Figure 6.3: Forward (a) and conjugate (b) reconstructions obtained in a system using a self-pumped phase conjugator.

Figure 6.4: Portion of the DHR modulator array, imaged through the signal beam path and captured by CCD #2.

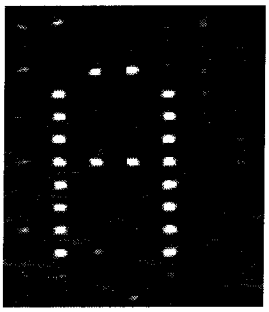


Figure 6.5: Forward reconstruction of a hologram of the DHR OEIC displaying the letter A.

fully recorded. For example, Figure 6.5 shows a reconstruction of a hologram of the DHR array displaying the Roman letter A.

6.2.4 Consequences of the shift between modulators and detectors in the OEIC

One of the benefits of conjugate readout is that the signal beam reconstructed during readout cycles self-focuses back on the OEIC. The wavefronts of the diffracted beam are identical to those of the corresponding signal beam during recording, but its direction of propagation is reversed. If a pattern such as those shown in Figures 6.4 and 6.5 is read out, its information-bearing areas[5] are focused back on the modulator electrodes/mirrors. However, the most compact and straightforward LCOS OEIC

[5]The regular grid of dots corresponding to modulator electrodes/mirrors on the OEIC.

designs require spatially separated detectors and modulators. For example, in the DHR-4 device, the detectors and modulators are offset by 61 μm (see Figure 3.9). In order for the OEIC to properly sense these areas, they should impinge on the detectors instead of the modulator electrodes/mirrors.

Fortunately, an easy solution exists to this problem. While the angular selectivity of holographic systems illustrated in Figures 6.1 and 6.2 is very strong in the plane of the Figures, it is much weaker in and out of this plane (see the discussion on page 170, including footnote 2). A small vertical tilt of the readout reference with respect to the recording reference will shift the conjugate reconstruction vertically on the surface of the OEIC with no appreciable loss of diffraction efficiency. If the OEIC is oriented in such a way that the shift between detectors and modulators is vertical, a minute vertical tilt of the conjugate reference allows the reconstructed data bits to be correctly sensed by the OEIC.

In the system employing a mirror loop (Figure 6.1), the desired tilt can be effected by slightly rotating one of the mirrors, for example the one at the bottom right corner of the loop.

Unfortunately, no such simple solution exists in the case of the system using a self-pumped phase conjugator. In this case, an additional two-position beam steerer could be inserted in the signal arm, between the OEIC and the recording material. This device would be programmed to slightly tilt the propagation axis of the reconstructed signal beam during readout cycles. The OEIC would have to be removed from the recording medium; the necessary image shift would build up over the distance between these two elements. The two-position beam steerer is easier to fabricate and less costly than the high-resolution units required for multiplexing.

Relatively simple schemes can be devised to allow each of the system architectures proposed in §5.5 to use OEICs with separated modulators and detectors[102].

6.3 Demonstration of sustained holograms using periodic refreshment

The experiments described in §6.2 were completed just before the Optical Society of America's 1995 Annual Meeting in Portland, Oregon, in September 1995, where the results were reported[76]. A few months later, Accuwave Corp. (Santa Monica, CA) demonstrated conjugate readout in a compact read-only holographic memory in which 50 wavelength-multiplexed holograms were recorded[77]. Since then, our group at Caltech performed experiments and demonstrations that expand on the original work in a number of ways:

Lens-less data path The new systems demonstrated lens-less conjugate readout. The earlier experimental setups shown in Figures 6.1 and 6.2 contained a 4-F imaging system between the OEIC and the recording medium to ease experimentation and measurements.

Better modulators New DHR devices were fabricated. Their modulators, assembled in our group's clean room in the new Moore Laboratory, possessed a considerably improved uniformity that allowed almost all the area of the device to be employed.

Captured conjugate reconstruction Conjugate holographic reconstructions were successfully captured, by slightly tilting a mirror as described in §6.2.4.

Better light efficiency In the new setups, beams were routed and split under electronic control, using liquid crystal cells and polarizing beam splitters. This led to a better utilization of the available laser power, and to faster recording and stronger holographic reconstructions.

Refreshment A hologram was sustained over 50 refresh/decay cycles using the DHR OEIC.

These recent experiments were performed with our colleague Ernest Chuang. We decided at the onset that we would generate the readout (conjugate) reference beam

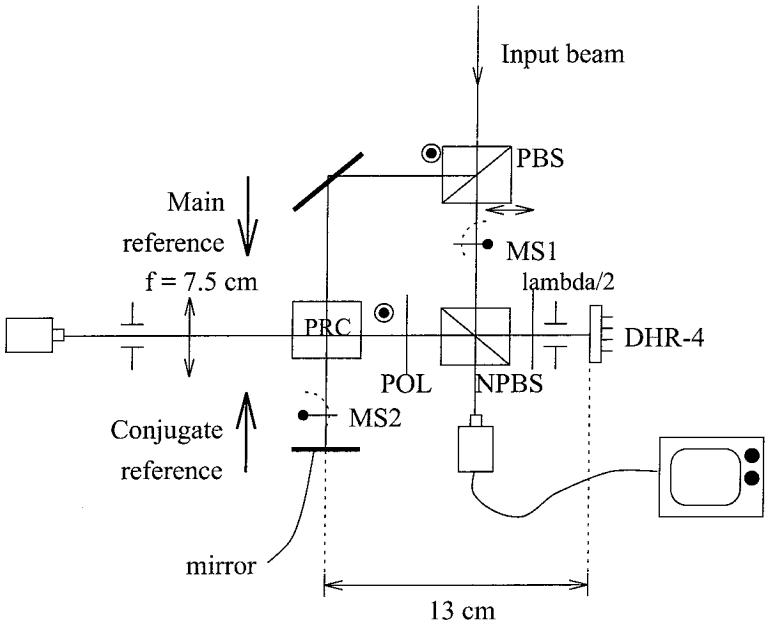


Figure 6.6: Conjugate readout demonstration system generating the readout reference using a single mirror. PRC = photorefractive crystal (recording medium), PBS = polarizing beam splitter, NPBS = non-polarizing beam splitter, MS = mechanical shutter, POL = polarizer.

using mirrors (as in Figures 6.1 and 5.3) instead of a phase conjugator (Figure 6.2), because of the issues and concerns enumerated on page 171.

6.3.1 Generation of the conjugate reference using a single mirror

We initially experimented briefly with the setup shown in Figure 6.6. This compact system is similar to that illustrated in Figure 5.3a. The input expanded beam (λ = 488 nm) enters the setup through the polarizing beam splitter (PBS) near the upper right corner of the schematic. The horizontally polarized beam transmitted by the PBS is the signal beam. It is reflected to the right by the non-polarizing beam splitter (NPBS), and modulated and reflected by the OEIC. The half-wave retardation plate and the polarizer ensure that the OEIC's modulators deliver their optimum contrast. The vertically polarized beam reflected to the left by the PBS becomes the main

Mode	MS1	MS2
recording	open	closed
fwd. readout	closed	closed
cjg. readout	closed	open

Table 6.1: State of the shutters in each operating mode of the system shown in Figure 6.6.

reference beam. It enters the recording medium ($LiNbO_3$:Fe) through its top face in the diagram of Figure 6.6. The relative strength of the signal and main reference beams can be adjusted by varying the polarization state of the input beam[6]. The components in the signal beam train to the left of the recording medium (lens, iris and camera) are only needed to observe forward ("normal") holographic reconstructions; they are not required for conjugate readout. Table 6.1 shows the state of the shutters in each operating mode of the system.

The mirror located below the recording medium in Figure 6.6 plays the same role as the mirror face of the recording medium in Figure 5.3a. During conjugate readout cycles, MS2 is opened, allowing the main reference to be reflected by the mirror, thereby providing the desired counter-propagating reference wave.

Holograms were successfully recorded using this system. Forward and conjugate reconstructions were obtained. However, the conjugate reconstructions were very weak, primarily because the conjugate reference beam was obtained from the main reference beam after the latter incurred a severe attenuation during its first pass through the thick (2 mm) recording medium. This problem was mentioned in §5.2.3. The ratio of the power diffracted in the forward and conjugate reconstructions was found to be as high as 38:1.

[6] A polarizing beam splitter transmits horizontally polarized light and reflects vertically polarized light.

6.3.2 System employing a mirror loop

Experimental setup

The system illustrated in Figure 6.7 is based upon the architecture shown in Figure 5.3c. Neither the forward nor the conjugate reference beam suffers unnecessary attenuation by traversing the recording medium more than once.

Holograms were recorded in a 30°-cut $BaTiO_3$ crystal, labeled PRC in Figure 6.7. $BaTiO_3$ was chosen for its recording speed. The input beam of wavelength 488 nm, controlled by mechanical shutter MS0, is split into a signal beam (going down) and a reference beam (going left) by polarizing beam splitter PBS1. The reference beam is steered to either the main reference path (directly to PRC) or the conjugate reference path (M1–M3, PRC) by the combined action of a 90° twisted nematic liquid crystal cell TNLC under computer control and a polarizing beam splitter PBS2. The signal beam is modulated and reflected toward the recording crystal by the combination of a polarizing beam splitter PBS3, a half-wave plate and the DHR integrated circuit. The forward and conjugate reconstructions were observed by using CCD cameras, a lens and a filtering iris; note that none of these elements are required for the normal operation of the module. A removable half-wave plate, located between the recording medium and PBS3, allowed the transmission of the conjugate reconstruction through PBS3, and its observation using CCD camera CCD2.

The two 90° twisted nematic liquid crystal modulators[131, 132], TNLC1 and TNLC2, provide electronic control of the path of the reference beam and of the relative strength of the signal and reference beams. TNLC1 and PBS1 are used to steer most of the input power to either the reference of the signal beam. TNLC1 rotates the polarization axis of the input beam by 90° during recording cycles. For the moment, ignore the half-wave phase retardation plate that follows it in the optical train. PBS1 transmits horizontally polarized light toward the OEIC and reflects vertically polarized light toward the reference path[7]. During recording operations, the signal and reference beams must both be strong. Most of the input power is sent in the

[7]See footnote 6 on page 179.

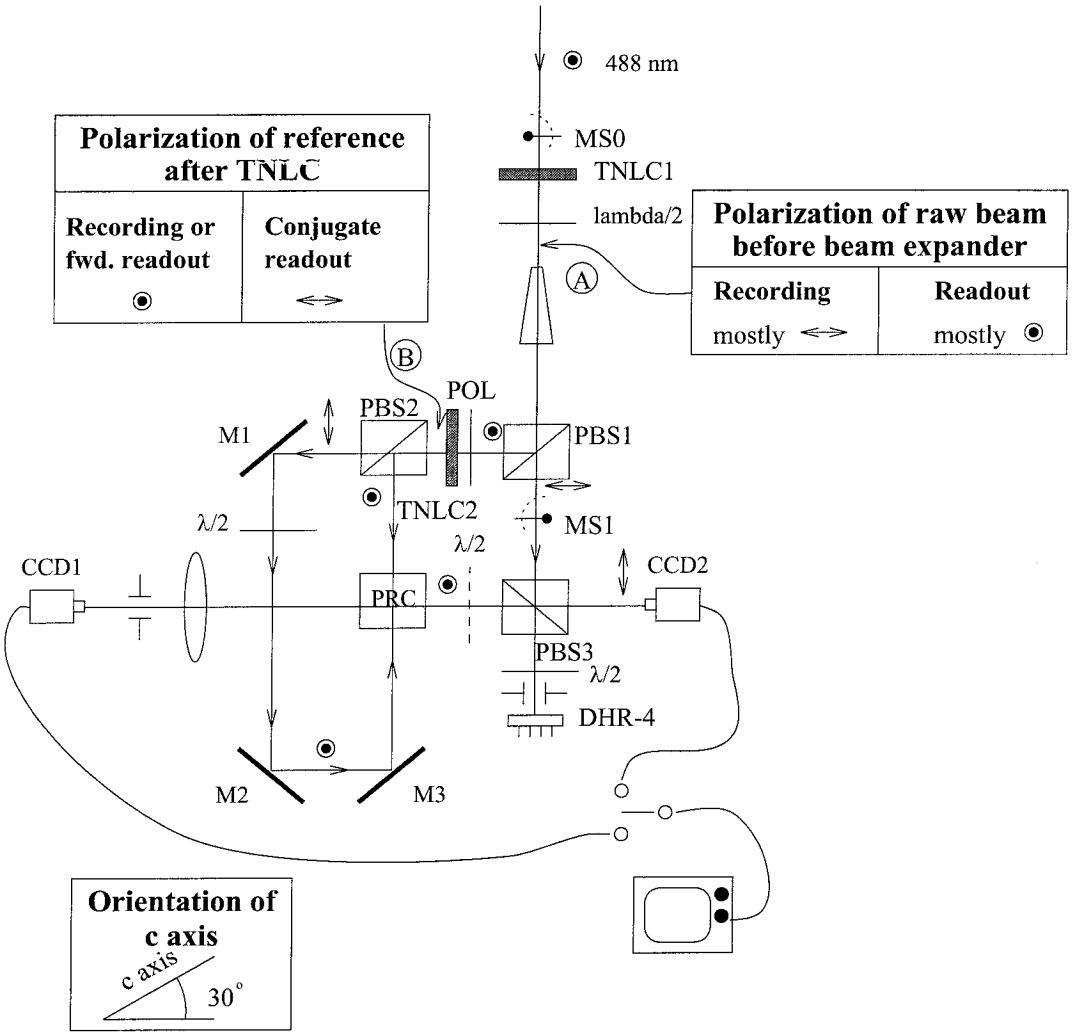


Figure 6.7: Conjugate readout demonstration system generating the readout reference using a mirror loop. PRC = photorefractive crystal (recording medium), PBS = polarizing beam splitter, NPBS = non-polarizing beam splitter, MS = mechanical shutter, POL = polarizer, M = mirror, CCD = charge-coupled device (camera), TNLC = twisted nematic liquid crystal.

signal arm, to avert the insertion loss of the DHR-4 OEIC (due to its low fill factor). This is accomplished by shorting the two electrodes of TNLC1, so that the full rotatory power of the cell's twisted structure is obtained and the beam is horizontally polarized at the output face of TNLC1. PBS1 and PBS3 transmit the horizontally polarized beam, which is then modulated by the OEIC. During readout cycles, the chip requires no illumination, and most of the input power should be directed to the appropriate reference beam, in order to ensure a strong holographic reconstruction. This is possible if the rotatory power of TNLC1 is suppressed by applying a relatively high voltage ($\approx$ 10 V) across its liquid crystal layer; the twisted structure is then destroyed, and the optic axis is parallel to the propagation axis of the input beam throughout the cell. As a result, the beam impinging on PBS1 is vertically polarized, and is (almost) fully reflected toward PBS2. The half-wave plate that follows TNLC1 in the optical train adds a bias angle to the orientation of the polarization axis at the output of TNLC1. When either the reference or the signal beam is selected to receive most of the input power, a fraction of the input power still goes to the other beam. The orientation of the half-wave retardation plate controls this fraction, which is small if the angle between either eigenaxis of the plate and the plane of Figure 6.7 is small.

The beam reflected by PBS1 becomes the reference beam. If PBS1 were an ideal polarizing beam splitter, this beam would be linearly and vertically polarized. An additional polarizer located between PBS1 and TNLC2 removes any horizontally polarized component of the beam resulting from imperfections in PBS1. If TNLC2 were absent or if its twisted liquid crystal conformation were destroyed by an applied voltage, the vertically polarized beam would be reflected by PBS2[8] to the recording reference path (down to the recording medium). With no applied voltage, TNLC2 rotates the polarization axis of the beam by 90°, and the beam is routed to the conjugate reference path (M1–M3, PRC). The half-wave plate located between M1 and M2 undoes the rotation of the polarization axis. The former routing is appropriate for recording cycles, in which the signal beam interferes with the recording reference; the latter routing is desirable in conjugate readout cycles, when the conjugate reference is

[8]See footnote 6 on page 179.

Mode	MS0	MS1	V_{TNLC1}[a]	Pol. at pt. A[b]	V_{TNLC2}[c]	Pol. at pt. B[d]
recording	open	open	0 V	mostly $\leftrightarrow$	10 V	$\odot$
fwd. readout	open	closed	10 V	mostly $\odot$	10 V	$\odot$
cjg. readout	open	closed	10 V	mostly $\odot$	0 V	$\leftrightarrow$

[a] Voltage across the liquid crystal layer of TNLC1.
[b] Polarization at point A.
[c] Voltage across the liquid crystal layer of TNLC2.
[d] Polarization at point B.

Table 6.2: State of the shutters, liquid crystal modulators and beam polarizations in the three operating modes of the system shown in Figure 6.7.

diffracted by the holographic gratings and produces a self-focusing reconstruction in the plane of the OEIC.

TNLC1 and TNLC2 are 90° twisted nematic liquid crystal cells fabricated in our laboratories at Caltech. Their substrates are glass plates coated with indium-tin-oxide and with uniaxially rubbed polyimide, which provides homogeneous alignment of the liquid crystal. The cell was assembled so that the rubbing axes of the two substrates were perpendicular. The plates are separated by a 10 μm gap, set by a mixture of glass fiber spacers and UV-cured adhesive in the periphery of the active area. The gap was filled with Merck's E7 nematic liquid crystal[90]. TNLC2, which routes the reference beam to the appropriate path, has a contrast ratio of more than 2000:1 in the normally-white mode (i.e., between crossed high-quality polarizers).

Table 6.2 summarizes the states of the shutters and liquid crystal modulators in each operating mode of the system.

The routing scheme described above results in a better light efficiency. The losses due to the non-polarizing beam splitter of Figure 6.1 are eliminated, and the electronically controlled dynamic routing scheme avoids wasting excessive power in the signal arm during readout. However, since M1–M3 and PBS2 do not form a Sagnac interferometer, the mirror loop cannot be aligned interferometrically. Alignment is performed by writing holograms and trying to detect them using the DHR-4 OEIC, making sure that a proper pixel registration is obtained.

Experimental results

A hologram of a binary pattern was recorded, allowed to decay and refreshed for 50 cycles under computer control. Specifically, after it reached the target conjugate diffraction efficiency, 1.2×10^{-4}, the hologram was exposed to the Bragg-matched main reference until its diffraction efficiency decayed to the refresh threshold, 7.0×10^{-5}. The hologram was then reconstructed using the conjugate reference and sensed and memorized by the DHR. The signal and main reference beams were then turned on, in order to strengthen the hologram until the diffraction efficiency was restored to its target value. The conjugate diffraction efficiency was 33% of the forward diffraction efficiency. Figure 6.8 shows the measured conjugate diffraction efficiency as a function of time. There were no errors at any cycle in any pixel of the detected data pages. Figure 6.9 shows the data page (the letters "CIT"), displayed on the DHR, imaged through the $BaTiO_3$ crystal and detected by CCD1 (a), as well as a forward reconstruction measured by CCD1 (b) and conjugate reconstructions measured by CCD2 after one (c) and fifty (d) cycles. Some non-uniform attenuation is visible in the forward reconstruction (Figure 6.9b). It is believed to be due to the attenuation of the reference beam as it propagates through the crystal (the total transmission through the 5.7 mm thick crystal is 12%). However, phase-conjugate readout (Figure 6.9c) compensates for this non-uniformity, since the reference beam enters the opposite face during readout. Table 6.3 lists the contrast ratio and signal-to-noise ratio (SNR) measured for each of these images, as well as the estimated probability of error. The signal to noise ratio is defined as $\mathrm{SNR} \equiv (\mu_1 - \mu_0) / \sqrt{\sigma_0^2 + \sigma_1^2}$, where μ_0 and μ_1 are the mean values of zero and one pixels, respectively, and σ_0^2 and σ_1^2 are the corresponding variances. Note that the conjugate reconstruction possesses a better SNR than its forward counterpart, and that the SNR does not degrade significantly as a result of refreshment.

The uniformity of the modulators is visibly better in Figure 6.9a than in Figure 6.4.

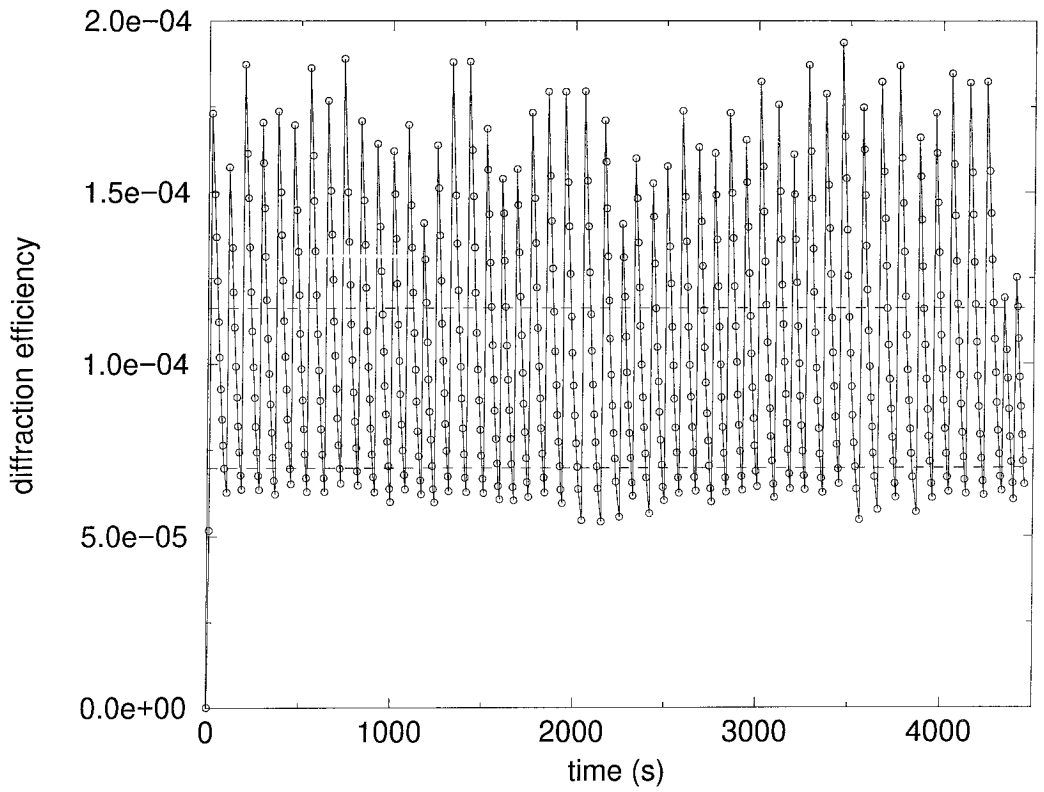


Figure 6.8: Evolution of the conjugate diffraction efficiency of a hologram during 50 refresh-decay cycles.

Image	CR	SNR	PE
DHR display	15.2:1	3.76:1	1.1×10^{-4}
fwd. rec.	18.8:1	2.24:1	2.2×10^{-3}
cjg. rec., 1 cyc.	11.0:1	3.23:1	6.9×10^{-4}
cjg. rec., 50 cyc.	11.7:1	3.03:1	1.0×10^{-3}

Table 6.3: Contrast ratio, signal to noise ratio and probability of error corresponding to the images shown in Figure 6.9.

Non-uniform pixel responses[9] restricted the usable area of the device[10] used in early experiments to approximately 9×5 pixels. The newer DHR device[11] was processed in the Psaltis group's clean room in the Moore Laboratory. The controlled environment of the clean room has noticeably increased yields and uniformity. One example of this improvement is the DHR OEIC shown in 6.9a, in which virtually all the pixels can be used simultaneously. We take this opportunity to reiterate our gratitude to Gordon and Betty Moore, whose gracious gift to Caltech made this clean room possible.

Subsequent experiment

Our colleague Ernest Chuang is now in charge of the continuation of the experiment described above. One of the early results of his work was the sustenance of three angular multiplexed holograms over 100 refresh/decay cycles[133]. Multiplexing was effected by rotating a motorized mount to which the recording medium was affixed.

[9]The non-uniformity is most pronounced at shorter wavelengths, which are unfortunately those at which the chosen recording media are sensitive.

[10]Sample 175, using modulators fabricated in the Steele Laboratory.

[11]Sample 191.

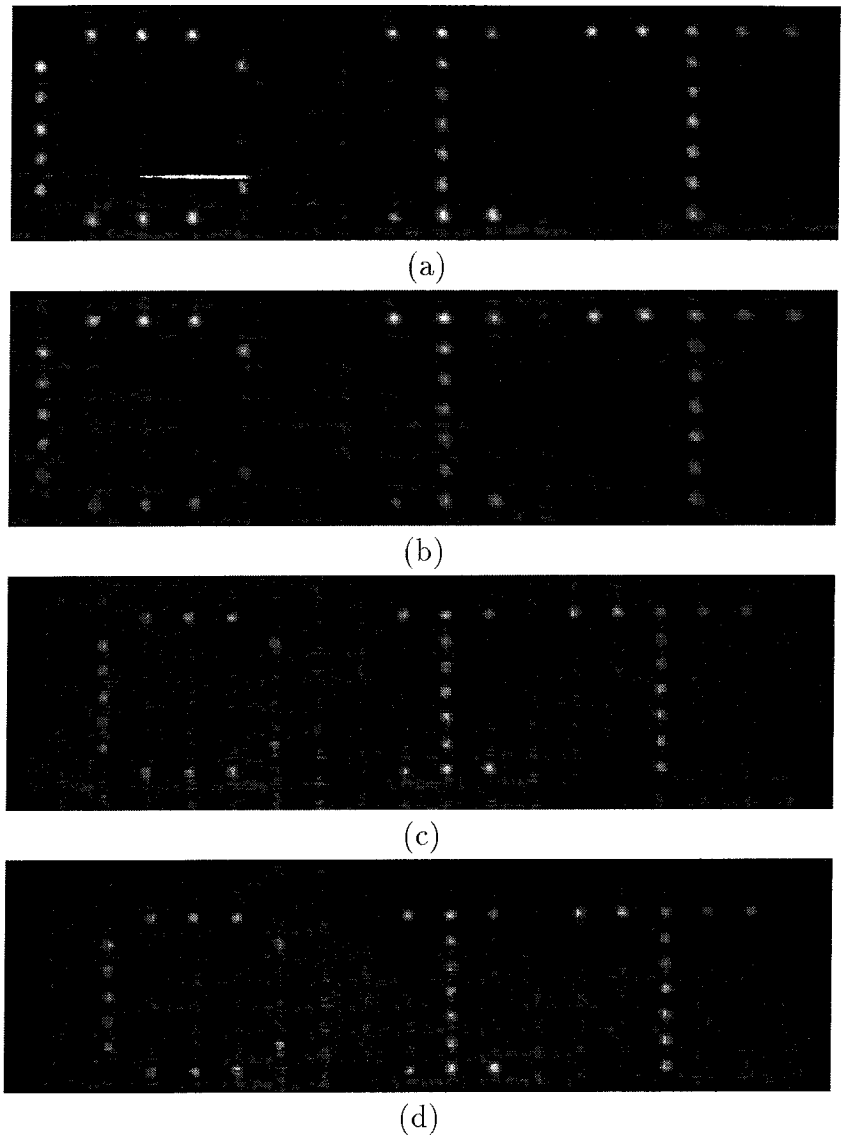


Figure 6.9: DHR display (a), forward reconstruction (b) and conjugate reconstructions after one (c) and fifty (d) refresh-decay cycles.

Bibliography

[1] Tom R. Halfhill. CDs for the gigabyte era. *BYTE*, 21(10):139–144, October 1996.

[2] Joseph W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill Physical and Quantum Electronics Series. McGraw-Hill, 1968.

[3] Pochi Yeh. *Introduction to Photorefractive Nonlinear Optics*. John Wiley & Sons, 1993.

[4] N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii. Holographic storage in electrooptic crystals. I. steady state. *Ferroelectrics*, 22:949–960, 1979.

[5] N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii. Holographic storage in electrooptic crystals. II. beam coupling — light amplification. *Ferroelectrics*, 22:961–964, 1979.

[6] Juan J. Amodei and David L. Staebler. Holographic pattern fixing in electrooptic crystals. *Applied Physics Letters*, 18(12):540–542, June 1971.

[7] L. Arizmendi, P. D. Townsend, M. Carrascosa, J. Baquedano, and J. M. Cabrera. Photorefractive fixing and related thermal effects in $LiNbO_3$. *Journal of Physics — Condensed Matter*, 3(28):5399–5406, 1991.

[8] Y. Qiao, S. Orlov, D. Psaltis, and R. R. Neurgaonkar. Electrical fixing of photorefractive holograms in $Sr_{0.75}Ba_{0.25}Nb_2O_6$. *Optics Letters*, 18(12):1004–1006, 1993.

[9] Kevin Curtis and Demetri Psaltis. Recording of multiple holograms in photopolymer films. *Applied Optics*, 31(35):7425–7428, December 1992.

[10] H. Lee, X.-G. Gu, and D. Psaltis. Volume holographic interconnections with maximal capacity and minimal cross talk. *Journal of Applied Physics*, 65(6):2191–2194, March 1989.

[11] A. B. Vander Lugt. Signal detection by complex spatial filtering. *IEEE Transactions on Information Theory*, 2(IT-10):139, 1964.

[12] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William J. Vetterling. *Numerical Recipes in C.* Cambridge University Press, 1988.

[13] Alan V. Oppenheim and Alan S. Willsky. *Signals and Systems.* Prentice-Hall, 1983.

[14] Demetri Psaltis and Michael Levene. Optical neural networks. In Su-Shing Chen and H. John Caulfield, editors, *Adaptive Computing: Mathematics, Electronics, and Optics*, volume CR55 of *Critical Reviews of Optical Science and Technology*, pages 141–149, Orlando, Florida, April 1994.

[15] Hsin-Yu Sidney Li. *Photorefractive 3-D Disks for Optical Data Storage and Artificial Neural Networks.* PhD thesis, California Institute of Technology, 1994.

[16] D. Psaltis, D. Brady, X. G. Gu, and S. Lin. Holography in artificial neural networks. *Nature*, 343(6256):325–330, 1990.

[17] John Hertz, Anders Krogh, and Richard G. Palmer. *Introduction to the Theory of Neural Computation.* Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, 1991.

[18] O. Matan, H. S. Baird, J. Bromly, C. J. C. Burges, J. S. Denker, L. D. Jackel, and Y. LeCun. Reading handwritten digits - a ZIP code recognition system. *Computer*, 25(7):59–63, July 1992.

[19] California Institute of Technology and London Business School. *Neural Networks in the Capital Markets*, Pasadena, California, November 1996.

[20] N. H. Farhat, D. Psaltis, A. Prata, and E. G. Paek. Optical implementation of the Hopfield model. *Applied Optics*, 24(10):1469–1475, 1985.

[21] Kelvin Wagner and Demetri Psaltis. Multilayer optical learning networks. *Applied Optics*, 26(23):5061–5076, 1987.

[22] Geoffrey W. Burr, Xin An, Fai H. Mok, and Demetri Psaltis. Large-scale rapid-access holographic memory. In *SPIE Topical Meeting on Optical Data Storage*, San Diego, 1995. Paper F2.3.

[23] Demetri Psaltis, David Brady, and Kelvin Wagner. Adaptive optical networks using photorefractive crystals. *Applied Optics*, 27(9):1752–1759, 1988.

[24] Hsin-Yu Li, Jean-Jacques P. Drolet, Demetri Psaltis, and Mark A. Handschy. Optoelectronic chip for the implementation of back error propagation. In *IEEE/LEOS Meeting, Santa Barbara, CA*, August 1992. Paper WA2.

[25] Amnon Yariv. *Optical Electronics.* Holt, Rinehart and Winston Series in Electrical Engineering. HRW-Saunders, fourth edition, 1991.

[26] C. A. Mead. Schottky barrier gate field-effect transistor. *Proc. IEEE*, 54:307, 1966.

[27] S. M. Sze. *Semiconductor Devices, Physics and Technology.* J. Wiley & Sons, 1985.

[28] S. M. Sze. *Physics of Semiconductor Devices.* John Wiley & Sons, second edition, 1981.

[29] Steven Lin, Annette Grot, Jiafu Luo, and Demetri Psaltis. GaAs optoelectronic neuron arrays. *Applied Optics*, 32(8):1275–1289, March 1993.

[30] A. C. Grot, D. Psaltis, K. V. Shenoy, and C. G. Fonstad. Integration of LEDs and GaAs circuits by MBE regrowth. *IEEE Photonics Technology Letters*, 6(7):819, July 1994.

[31] K. V. Shenoy, C. G. Fonstad, A. C. Grot, and D. Psaltis. Monolithic optoelectronic circuit — design and fabrication by epitaxial growth on commercial VLSI GaAs MESFET's. *IEEE Photonics Technology Letters*, 7(5):508, May 1995.

[32] Jiafu Luo, Annette Grot, and Demetri Psaltis. High-responsivity optical FETs fabricated on a FET-SEED structure. *IEEE Photonics Technology Letters*, 7(7):760–762, July 1995.

[33] K. W. Goossen, J. A. Walker, L. A. D'Asaro, S. P. Hui, B. Tseng, R. Leibenguth, D. Kossives, D. D. Bacon, D. Dahringer, L. M. F. Chirovsky, A. L. Lentine, and D. A. B. Miller. GaAs MQW modulators integrated with silicon CMOS. *IEEE Photonics Technology Letters*, 7(4):360–362, April 1995.

[34] Paul M. Alt and Peter Pleshko. Scanning limitations of liquid-crystal displays. *IEEE Transactions on Electron Devices*, ED-21(2):146–155, February 1974.

[35] Jürgen Nehring and Allan R. Kmetz. Ultimate limits for matrix addressing of RMS-responding liquid-crystal displays. *IEEE Transactions on Electron Devices*, ED-26(5):795–801, May 1979.

[36] S. Mendis, S. E. Kemeny, and E. R. Fossum. CMOS active pixel image sensor. *IEEE Transactions on Electron Devices*, 41(3):452–453, March 1994.

[37] Eric R. Fossum, Sunetra Mendis, and Sabrina Kemeny. Active-pixel sensor with intra-pixel charge transfer, November 1995. U.S. Patent 5,471,515.

[38] J. Nakamura, S. E. Kemeny, and E. R. Fossum. CMOS active pixel image sensor with simple floating gate pixels. *IEEE Transactions on Electron Devices*, 42(9):1693–1694, September 1995.

[39] Sunetra K. Mendis, Bedabrata Pain, Robert H. Nixon, and Eric R. Fossum. Design of a low-light-level image sensor with on-chip sigma-delta analog-to-digital conversion. *Proceedings of the SPIE*, 1900:31–39, 1993.

[40] Eric R. Fossum. CMOS image sensors: Electronic camera on a chip. In *IEEE International Electron Devices Meeting Technical Digest*, Washington, D.C., 1995.

[41] Fai H. Mok, Geoffrey W. Burr, and Demetri Psaltis. System metric for holographic memory systems. *Optics Letters*, 21(12):896–898, June 1996.

[42] Fai H. Mok. private communication.

[43] G. D. Bacher, M. P. Chiao, G. J. Dunning, M. B. Klein, C. C. Nelson, and B. A. Wechsler. Ultralong dark decay measurements in $BaTiO_3$. *Optics Letters*, 21(1):18–20, January 1996.

[44] D. Psaltis, F. Mok, and H. Y. S. Li. Non-volatile storage in photorefractive crystals. *Optics letters*, 19(3):210–212, 1994.

[45] K. Buse, L. Holtmann, and E. Krätzig. Activation of $BaTiO_3$ for infrared holographic recording. *Laser Focus World*, 2–3(85):183–186, 1991.

[46] Y. Qiao and D. Psaltis. Sampled dynamic holographic memory. *Optics Letters*, 17(19):1376–1378, 1992.

[47] S. Ducharme, J. C. Scott, R. J. Twieg, and W. E. Moerner. Observation of the photorefractive effect in a polymer. *Physical Review Letters*, 66(14):1846–1849, 1991.

[48] K. Meerholz, B. L. Volodin, B. Sandalphon Kippelin, and N. Peyghambarian. A photorefractive polymer with high optical gain and diffraction efficiency near 100%. *Nature*, 371(6497):497–500, October 1994.

[49] George G. Malliaras. *Photorefractivity in Polymers.* PhD thesis, University of Groningen, Groningen, The Netherlands, November 1995.

[50] B. Kippelen, K. Meerholz, B. L. Sandalphon Volodin, and N. Peyghambarian. Photorefractive polymers and their applications. *Molecular Crystals and Liquid Crystals, Science and Technology, Section A*, 283:109–114, 1996.

[51] J. J. Stankus, S. M. Silence, W. E. Moerner, and G. C. Bjorklund. Electric-field-switchable stratified volume holograms in photorefractive polymers. *Optics Letters*, 19(18):1480–1482, September 1994.

[52] U. S. Rhee, H. J. Caulfield, J. Shamir, C. S. Vikram, and M. M. Mirsalehi. Characteristics of the DuPont photopolymer for angularly multiplexed page-oriented holographic memories. *Optical Engineering*, 32(8):1839–1847, August 1993.

[53] Kevin Curtis and Demetri Psaltis. Characterization of the DuPont photopolymer for three-dimensional holographic storage. *Applied Optics*, 33(23):5396–5399, August 1994.

[54] Allen Pu and Demetri Psaltis. High-density recording in photopolymer-based holographic three-dimensional disks. *Applied Optics*, 35(14):2389–2398, May 1996.

[55] R. V. Johnson and Jr. A. R. Tanguay. Stratified volume holographic optical elements. *Optics Letters*, 13(3):189–191, March 1988.

[56] G. P. Nordin and Jr. A. R. Tanguay. Photopolymer-based stratified volume holographic optical elements. *Optics Letters*, 17(23):1709–1711, December 1992.

[57] R. De Vré and L. Hesselink. Analysis of stratified volume holographic optical elements. *Journal of the Optical Society of America B*, 11(9):1800–1808, 1994.

[58] M. P. Bernal, H. Coufal, J. A. Hoffnagle, C. M. Jefferson, R. M. Macfarlane, R. M. Shelby, G. T. Sincerbox, and P. Wimmer *et al.* A precision tester for studies of holographic optical storage materials and recording physics. *Applied Optics*, 35(14):2360–2374, May 1996.

[59] IBM Holographic Optical Storage Team. Holographic storage promises high data density. *Laser Focus World*, 32(11):81–93, November 1996.

[60] Frank Goodenough. IEDM showcases tomorrow's analog semiconductors. *Electronic Design*, 44(25):42–52, December 1996.

[61] Paul McGoldrick. Laser increases possible DVD capacity threefold. *Electronic Design*, 44(25):27, December 1996.

[62] Rick De Meis. InGaN diode laser emits blue-violet light. *Laser Focus World*, 32(4):18, April 1996.

[63] J. L. Jewell, J. P. Harbison, and A. Scherer. Microlasers. *Scientific American*, 265(5):56–62, 1991.

[64] Antti Vasara, Mohammad R. Taghizadeh, Jari Turunen, Jan Westerholm, Eero Noponen, Hiroyuki Ichikawa, J. Michael Miller, Tommi Jaakkola, and Sirpa Kuisma. Binary surface-relief gratings for array illumination in digital optics. *Applied Optics*, 31(17):3320–3336, June 1992.

[65] Patrick F. Brinkley, Stephen T. Kowel, and Chinghua Chu. Liquid crystal adaptive lens: Beam translation and field meshing. *Applied Optics*, 27(21):4578–4586, November 1988.

[66] N. A. Riza and M. C. Dejule. 3-terminal adaptive nematic liquid-crystal lens device. *Optics Letters*, 19(14):1013–1015, July 1994.

[67] Terry A. Dorschner and Daniel P. Resler. Optical beam steerer having subaperture addressing, March 1992. U.S. Patent 5,093,740.

[68] Paul F. McManamon, Edward A. Watson, Terry A. Dorschner, and Lawrence J. Barnes. Applications look at the use of liquid crystal writable gratings for steering passive radiation. *Optical Engineering*, 32(11):2657–2664, November 1993.

[69] D. P. Resler, D. S. Hobbs, R. C. Sharp, L. J. Friedman, and T. A. Dorschner. High-efficiency liquid-crystal optical phased-array beam steering. *Optics Letters*, 21(9):689–691, May 1996.

[70] Fai Ho Mok. Angle-multiplexed storage of 5000 holograms in lithium niobate. *Optics Letters*, 18(11):915–917, 1993.

[71] Geoffrey William Burr, Fai Ho Mok, and Demetri Psaltis. Spatial and angle multiplexing using the long interaction length architecture. In *Optical Society of America Annual Meeting*, October 1993. Paper TuH6.

[72] Geoffrey William Burr, Fai Ho Mok, and Demetri Psaltis. Large scale volume holographic storage in the long interaction length architecture. In *Proceedings of the SPIE*, 1994.

[73] Xin An and Demetri Psaltis. Experimental characterization of an angle-multiplexed holographic memory. *Optics Letters*, 20(18):1913–1915, September 1995.

[74] Hsin-Yu Sidney Li, Yong Qiao, and Demetri Psaltis. Optical network for real-time face recognition. *Applied Optics*, 32(26):5026–5035, September 1993.

[75] Jack Feinberg. Self pumped continuous wave phase conjugator using internal reflection. *Optics Letters*, 7:486, 1982.

[76] Jean-Jacques P. Drolet, George Barbastathis, Jay S. Patel, and Demetri Psaltis. Liquid crystal devices for volume holographic memories. In *OSA Annual Meeting*, September 1995.

[77] Feng Zhao and Koichi Sayano. Compact read-only memory with lensless phase-conjugate holograms. *Optics Letters*, 21(16):1295–1297, August 1996.

[78] MOSIS, Information Sciences Institute, University of Southern California, Marina del Ray, California. *MOSIS User Guide*, release 3.1 edition, August 1992.

[79] Jean-Jacques P. Drolet, Jay S. Patel, Konstantinos G. Haritos, Weihua Xu, Axel Scherer, and Demetri Psaltis. Hybrid-aligned nematic liquid-crystal modulators fabricated on VLSI circuits. *Optics Letters*, 20(21):2222–2224, November 1995.

[80] Kristina M. Johnson, Douglas J. McKnight, and Ian Underwood. Smart spatial light modulators using liquid crystals on silicon. *IEEE Journal of Quantum Electronics*, 29(2):699–714, February 1993.

[81] Ian Underwood, David G. Vaas, Antony O'Hara, Dwayne C. Burns, Peter W. McOwan, and James Gourlay. Improving the performance of liquid-crystal-over-silicon spatial light modulators: Issues and achievements. *Applied Optics*, 33(14):2768–2774, May 1994.

[82] Douglas J. McKnight, Mark A. Follett, and Kristina M. Johnson. Liquid crystal over silicon spatial light modulators. In *Opt. Comput. Int. Conf.*, Edinburg, August 1994.

[83] Douglas J. McKnight, David G. Vass, and Richard M. Sillitto. Development of a spatial light modulator: a randomly addressed liquid-crystal-over-nMOS array. *Applied Optics*, 28(22):4757–4762, November 1989.

[84] Noel A. Clark and Sven T. Lagerwall. Submicrosecond bistable electro-optic switching in liquid crystals. *Applied Physics Letters*, 36(11):899–901, June 1980.

[85] L. K. Cotter, T. J. Drabik, R. J. Dillon, and M. A. Handschy. Ferroelectric-liquid-crystal/silicon-integrated-circuit spatial light modulator. *Optics Letters*, 15(5):291–293, March 1990.

[86] Timothy J. Drabik and Mark A. Handschy. Silicon VLSI/ferroelectric liquid crystal technology for micropower optoelectronic computing devices. *Applied Optics*, 29(35):5220–5223, December 1990.

[87] David A. Jared and Kristina M. Johnson. Optically addressed thresholding very-large-scale-integration/liquid crystal spatial light modulators. *Optics Letters*, 16(12):967–969, June 1991.

[88] Douglas J. McKnight, Kristina M. Johnson, and Roylnn A. Serati. 256×256 liquid-crystal-on-silicon spatial light modulator. *Applied Optics*, 33(14):2775–2784, May 1994.

[89] R. Barberi and G. Barbero. Variational calculus and simple applications of continuum theory. In Iam-Choon Khoo and Francesco Simoni, editors, *Physics of Liquid Crystalline Materials*, chapter IX. Gordon and Breach Science Publishers, 1988.

[90] E. Merck, Industrial Chemical Division. Liquid crystal mixtures for electro-optic displays, September 1992.

[91] Axel Scherer, Jean-Jacques P. Drolet, and Demetri Psaltis. Alignment of liquid crystals on porous materials. U.S. patent pending (serial no. 08/612,382), 1995.

[92] Ching-Fa Yeh, Juing-Yi Cheng, and Jau-Hone Lu. The characterization of Al_2O_3 prepared by anodic oxidation. *Japanese Journal of Applied Physics*, 32(6A):2803–2808, 1993.

[93] K. Shimizu, K. Kobayashi, G. E. Thompson, and G. C. Wood. Development of porous anodic films on aluminum. *Philosophical Magazine A*, 66(4):643–652, 1992.

[94] R. Koëtz, B. Schnyder, and C. Barbero. Anodic oxidation of aluminum in sulphuric acid monitored by *ex-situ* and *in-situ* spectroscopic ellipsometry. *Thin Solid Films*, 233:63–68, 1993.

[95] V. P. Parkhutik and V. I. Shershulsky. Theoretical modelling of porous oxide growth on aluminum. *Journal of Physics D*, 25(8):1258–1263, 1992.

[96] G. Patermarakis, P. Lenas, Ch. Karavassilis, and G. Papayiannis. Kinetics of growth of porous anodic Al_2O_3 films on Al metal. *Electrochimica Acta*, 36(3):709–725, 1991.

[97] Lev Mikhaĭlovich Blinov and Vladimir G. Chigrinov. *Electrooptic Effects in Liquid Crystals*. Springer-Verlag, 1994.

[98] Carver A. Mead. *Analog VLSI and Neural Systems*. Addison-Wesley, 1989.

[99] Francisco H. De La Moneda, Eugene R. Chenette, and Aldert van der Ziel. Noise in phototransistors. *IEEE Transactions on Electron Devices*, 18(6):340–346, June 1971.

[100] T. Delbrück and C. A. Mead. Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits. Technical report, California Institute of Technology, Computation and Neural Systems Program, Pasadena, California, February 1995. CNS Memo No. 30.

[101] Janeen D. W. Anderson and Carver A. Mead. MOS device for long-term learning, September 1990. U.S. Patent 4,953,928.

[102] Demetri Psaltis, Jean-Jacques P. Drolet, and George Barbastathis. Compact architecture for volume holographic memories. U.S. Patent pending; application filed in September 1996.

[103] M. White, D. Lampe, F. Blaha, and I. Mack. Characterization of surface channel CCD image arrays at low light levels. *IEEE Journal of Solid-State Circuits*, 9:1–13, 1974.

[104] Phillip E. Allen and Douglas R. Holberg. *CMOS Analog Circuit Design.* Holt, Rinehart and Winston, 1987.

[105] Paul R. Gray and Robert G. Meyer. *Analysis and Design of Analog Integrated Circuits.* John Wiley & Sons, second edition, 1984.

[106] Jiafu Luo. *Monolithic GaAs VLSI Optoelectronic Neuron Arrays.* PhD thesis, California Institute of Technology, 1996.

[107] D. A. B. Miller, D. S. Chemla, T. C. Damen, T. H. Wood, C. A. Burrus, A. C. Gossard, and W. Wiegmann. The quantum well self-electrooptic effect device — optoelectronic bistability and oscillation, and self-linearized modulation. *IEEE Journal of Quantum Electronics*, 21(9):1462–1476, 1985.

[108] A. V. Krishnamoorthy, J. E. Ford, K. W. Goossen, J. A. Walker, A. L. Lentine, L. A. D'Asaro, S. P. Hui, B. Tseng, R. Leibenguth, D. Kossives, D. Dahringer, L. M. F. Chirovsky, F. E. Kiamilev, G. F. Aplin, R. G. Rozier, and D. A. B. Miller. Implementation of a photonic page buffer based on GaAs MQW modulators bonded directly over active silicon VLSI circuits. In *Proceedings of the OSA Topical Meetings on Optical Computing*, Salt Lake City, March 1995.

[109] A. V. Krishnamoorthy, A. L. Lentine, K. W. Goossen, J. A. Walker, T. K. Woodward, J. E. Ford, G. F. Aplin, L. A. D'Asaro, S. P. Hui, B. Tseng, R. Leibenguth, D. Kossives, D. Dahringer, L. M. F. Chirovsky, and D. A. B. Miller. 3-D integration of MQW modulators over active sub-micron CMOS circuits: 375 Mb/s transimpedance receiver-transmitter circuit. *IEEE Photonics Technology Letters*, 7(11):1288–1290, November 1995.

[110] Tobias Delbrück. Bump circuits for computing similarity and dissimilarity of analog voltages. CNS Memo 10, California Institute of Technology, Pasadena, California 91125, May 1991.

[111] National technology roadmap for semiconductors. Semiconductor Industry Association report published by Sematech (Austin, TX), 1994.

[112] E. Jakeman and E. P. Raynes. Electro-optic response times in liquid crystals. *Physics Letters*, 39A(1):69–70, April 1972.

[113] Kevin Curtis. *3-D Photopolymer Disks for Correlation and Data Storage, and Cross-Talk in Volume Holographic Memories.* PhD thesis, California Institute of Technology, 1994.

[114] Micron Technology. MT4LC16M4G3/H9 16 MEG $\times$ 4 DRAM preliminary data sheet, March 1997.

[115] Maury Wright. Disk drives at 40: Lean, mean storage machines. *EDN*, 41(23):51–64, November 1996.

[116] Mark A. Neifeld and Mark McDonald. Error correction for increasing the usable capacity of photorefractive memories. *Optics Letters*, 19(18):1483–1485, September 1994.

[117] M. A. Neifeld and J. D. Hayes. Error-correction schemes for volume optical memories. *Applied Optics*, 34(35):8183–8191, December 1995.

[118] Alfred B. Vollmer. World's first 1-Gbit DRAM announced at Electronica. *Electronic Design*, 45(1):28, January 1997.

[119] E. Adler, J. K. Debrosse, S. F. Geissler, S. J. Holmes, M. D. Jaffe, J. B. Johnson, C. W. Koburger, J. B. Lasky, B. Lloyd, G. L. Miles, J. S. Nakos, W. P. Noble, S. H. Voldman, M. Armacost, and R. Ferguson. The evolution of IBM CMOS DRAM technology. *IBM Journal of Research and Development*, 1995.

[120] N. Kasai, M. Sakao, T. Ishijima, E. Ikawa, H. Watanabe, T. Takeshima, N. Tanabe, K. Terada, and T. Kikkawa. A capacitor over bit-line (COB) stacked capacitor cell using local interconnect layer for 64 Mb DRAMs. *IEICE Transactions on Electronics*, E76C(4):548–555, 1993.

[121] J. P. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead. Winner-take-all networks of $o(n)$ complexity. Technical Report Caltech-CS-TR-21-88, California Institute of Technology, Computer Science Department, 1989.

[122] Kelvin Wagner and Timothy M. Slagle. Optical competitive learning with VLSI/liquid-crystal winner-take-all modulators. *Applied Optics*, 32(8):1408–1435, March 1993.

[123] G. A. Rakuljic, V. Leyva, and A. Yariv. Optical data storage by using orthogonal wavelength-multiplexed volume holograms. *Optics Letters*, 17(20):1471–1473, October 1992.

[124] C. Denz, G. Pauliat, G. Roosen, and T. Tschudi. Volume hologram multiplexing using a deterministic phase encoding method. *Optics Communications*, 85(2–3):171–176, 1991.

[125] C. Denz, G. Pauliat, G. Roosen, and T. Tschudi. Potentialities and limitations of hologram multiplexing by using the phase-encoding technique. *Applied Optics*, 31(26):5700–5705, September 1992.

[126] C. Alves, G. Pauliat, and G. Roosen. Dynamic phase-encoding storage of 64 images in a $BaTiO_3$ photorefractive crystal. *Optics Letters*, 19(22):1894–1896, November 1994.

[127] George Barbastathis, Michael Levene, and Demetri Psaltis. Shift multiplexing with spherical reference waves. *Applied Optics*, 35(14):2403–2417, May 1996.

[128] Jean-Jacques P. Drolet, George Barbastathis, and Demetri Psaltis. Integrated optoelectronic interconnects using liquid-crystal-on-silicon VLSI. In *SPIE Proceedings CR62: Optoelectronic Interconnects and Packaging*, pages 106–131, January 1996.

[129] Jean-Jacques P. Drolet, Ernest Chuang, George Barbastathis, and Demetri Psaltis. Compact integrated dynamic holographic memory with refreshed holograms. *Optics Letters*, 1997. accepted for publication.

[130] Eugene Hecht. *Optics*. Addison-Wesley, second edition, 1974.

[131] C. H. Gooch and H. A. Tarry. Optical characteristics of twisted nematic liquid-crystal films. *Electronics Letters*, 10(1):1–4, January 1974.

[132] Amnon Yariv and Pochi Yeh. *Optical Waves in Crystals*, pages 143–147. Wiley, 1984.

[133] Ernest Chuang, Jean-Jacques P. Drolet, George Barbastathis, and Demetri Psaltis. Compact architecture for holographic storage. In *LEOS'96*, Boston, 1996. Paper TuD1 (invited).

Glossary and list of abbreviations

active pixel A pixel containing at least part of an amplifier.

B byte, or 8 bits.

b bit.

CCD Charge-coupled device. A type of analog shift register consisting of a long MOSFET with many gates. Each gate controls an underlying potential well in which charge can be accumulated. CCDs are often employed as optical detector arrays; the charges accumulated under the gates during a well defined integration time are shifted to a sense amplifier.

CD Compact disc.

CD-ROM Compact disc read-only memory.

CMOS Complementary metal-oxide-semiconductor. A CMOS integrated circuit can contain both p- and n-channel MOSFETs.

DHR Dynamic hologram refresher. An optoelectronic integrated circuit with pixels containing an optical input (photodetector), an optical output (modulator or emitter) and a memory element. The device can refresh data pages that slowly decay in a read/write holographic memory.

director A unit vector pointing in the locally averaged direction of the long axis of the oblong liquid crystal molecules.

DRAM Dynamic RAM. A type of RAM in which data is transiently stored as a charge on a capacitor. The data stored in DRAM must be frequently refreshed. The refresh periods are typically 32 ms to 128 ms as of this writing.

DVD A standard for optical discs offering storage capacities of up to 17 GB.[12]

FLC Ferroelectric liquid crystal. A class of liquid crystals that exhibit a permanent electric dipole and that are employed for high-performance light modulation.

HAN Hybrid-aligned nematic. A type of liquid crystal modulators in which one substrate induces homogeneous alignment while the other imposes homeotropic alignment.

HAN-on-VLSI Hybrid-aligned nematic (modulators) fabricated on VLSI devices. This is a type of LCOS modulators.

IC Integrated circuit.

LCOS Liquid crystal on silicon.

LED Light emitting diode.

MBE Molecular beam epitaxy.

MIS Metal-insulator-semiconductor. The MOS diode is an MIS structure in which the insulator is an oxidized layer of semiconductor.

MOS Metal-oxide-semiconductor. A structure consisting of a thin ($\sim$ 10 Å to $\sim$ 500 Å) layer of silicon dioxide comprised between a highly conductive sheet (metal or highly doped polycrystalline silicon) and a semiconductor bulk.

MOSFET MOS field-effect transistor. A transistor consisting of an MOS structure with two adjacent diffusions. The highly conductive layer of the MOS structure is called the *gate* of the transistor; it controls a *channel* between the two diffusions (called *source* and *drain*) in the underlying semiconductor bulk.

MQW Multiple quantum well.

[12]DVD is said to stand for digital videodisc or digital versatile disc.

neuromorphic system A system that borrows key features from biological neural systems, such as massively parallel densely connected networking of simple processors.

NTRS National Technology Roadmap for Semiconductors[111].

OEIC Optoelectronic integrated circuit.

PBS Polarizing beam splitter.

poly *See* polysilicon.

polysilicon Polycrystalline silicon. Used as a gate material in most CMOS processes. Also employed as an interconnection layer and to form linear resistors.

RAM Random-access memory. Refers to read/write semiconductor memory, generally with short and approximately equal read and write cycle times. RAM is usually volatile; its contents are lost when power is removed. The acronym is misleading because other types of memory, including ROM, also feature random-access capability.

ROM Read-only memory.

SEED Self-electrooptic effect device.

SIA Semiconductor Industry Association.

SLM Spatial light modulator. A device for imprinting a two-dimensional modulation on a light beam.

VLSI Very large scale integration.

Index