phys. stat. sol. (b) 150, 447 (1988)

Subject classification: 78.20

Universität Erlangen-Nürnberg[1]) *(a) and AT & T Bell Laboratories, Holmdel*[2]) *(b)*

Digital Optics

Architecture and Systems Requirements

By

N. Streibl (a) and M. E. Prise (b)

Digital optics is a technology for (i) processing, (ii) transport, and (iii) storage of optical digital information. Digital optics offers both, the high temporal bandwidth known from fiber-communications, as well as the high connectivity and information density of optical imaging. Digital optical circuits may be constructed by cascading two-dimensional planar arrays of optical logic gates. The connection between these arrays is performed by light beams going through the free space above a chip. The system design and the architecture have to be adapted to optics. Some requirements for optical systems and for optical logic gates will be discussed.

Digitale Optik ist eine Technologie (i) zur Verarbeitung, (ii) zum Transport und (iii) zur Speicherung optisch-digitaler Informationen. Digitale Optik bietet sowohl die hohe Zeitbandbreite, die man aus der Glasfaser-Kommunikation kennt, als auch die hohe „Konnektivität" und Informationsdichte der optischen Abbildung. Digitale optische Schaltkreise können aufgebaut werden durch Kaskadieren planarer zweidimensionaler Matrizen von optisch-logischen Gattern. Die Verbindungen zwischen diesen Schaltmatrizen erfolgen durch Lichtstrahlen, die durch den freien Raum über dem Chip verlaufen. Der System-Entwurf und die Architektur müssen an die Optik angepaßt sein. Einige Anforderungen an die optische Systeme sowie an die optischen logischen Gatter werden diskutiert.

1. Digital Optics

Today optical data communication through fibers is in general use. Optical mass storage has become a common technology. For data processing, however, the optical signals have to be converted into electrical signals so far. Today, digital optical processing is a challenging research problem.

Digital optics is a comprehensive technology for (i) data processing, (ii) data transport, and (iii) information storage.

The key components for digital optical data processing are (i) two-dimensional opto-electronic or opto-optical logic gate arrays, (ii) modules for parallel optical interconnections through free space and for beam shaping and (iii) system design and architectures adapted to optics.

The key technologies are (i) fabrication and processing of semiconductor materials for the optical logic gate arrays, (ii) classical optics, micro-optics and holography for interconnections and beam shaping and (iii) laser technology for the optical power supply of the gate arrays.

The connectivity (i.e. the number of independent data channels per chip) is 10^4 to 10^6 in optics. Electronic integrated circuits have a connectivity on the order of a few

---

[1]) Erwin-Rommel-Str. 1, D-8520 Erlangen, FRG.

[2]) Holmdel, New Jersey 07733, USA.

hundred. Free-space optical interconnects by imaging systems allow parallel access to two-dimensional arrays of data.

Integrated optics and fiber communications already offer high temporal bandwidth and a certain amount of information processing capability: Waveguides are good for long-distance interconnections at high data rates with low attenuation. Typically they are used for point-to-point interconnects. In the context of long-distance communication it is mostly sufficient to interconnect only a few participants or terminals. On the other hand, in a data processing system we find mainly moderate- and short-distance interconnections. A large number of these interconnections is necessary for complex digital circuits. Therefore, free-space chip-to-chip array interconnections by imaging systems or by holographic optical elements are more suitable than waveguides for this application. Waveguide-based switching devices or logical gates have been investigated in Integrated Optics. They are sometimes ultrafast, but usually very long in the propagation direction, in order to increase the interaction length. Thus, huge areas on a chip are required for moderately complex circuits and large scale integration of optical logic becomes virtually impossible. Hence, waveguide switches may find an application in ultrafast circuits with a low degree of complexity. In contrast, optical devices based on III/V semiconductors are highly integrable and they are easily interfaced to electronics, which might prove advantageous, especially now, in the early phase of digital optics.

Hence, in the future, high performance digital circuits will make use of optical interconnections because of two main reasons: (i) the large temporal bandwidth of optics which allow high data rates and (ii) the large connectivity of optics which allows parallel communications. The energy for optical switching and logic is presently comparable with electronics. We will discuss the minimum requirements for optical logic gates and for the optical systems to be useful in digital optics.

2. Architecture and Systems Design

Ultrafast transistors with switching times of a few picoseconds have been reported. The system time constants for complex electronic digital circuits are, however, two orders of magnitude slower than those of the fastest transistors. The reason is the difficulty in implementing large numbers of high-bandwidth electrical interconnections: They act as microwave emitters and they need to be terminated correctly. In addition, the interconnection lengths have to be controlled very accurately, in order not to introduce clock skews.

Free-space array interconnections by imaging systems or by holographic optical elements [1 to 6] have been proposed to send data to or take data out from a chip through the free space above the chip. "Space sharing" of beams from different optical logic gates is possible, if the interconnections are regular (or "space-invariant"). Mathematically, such space-invariant interconnect patterns can be described by convolutions. Physically, they can be implemented by slightly modified imaging systems (e.g. spatial filtering). The price for deviating from strict space-invariance is optical switches with large active area, low packing density, and/or complicated, non-synchronous optical interconnect systems. The reason is: Optical switches with low energy dissipation are necessarily small. Therefore, large apertures are necessary for addressing them. For a space-variant (or random-) interconnection some sort of multiplexing is required. Either space (= packing density) or aperture space (= resolution, which translates into energy dissipation) have to be traded for the irregularity of the connections [7, 8]. Wavelength multiplexing cannot be used with resonant devices, and for time multiplexing we do not have appropriate devices either. Polarization offers only a multiplex rate of two. Hence the use of regular interconnections is mandatory for optical data processing.

Since regular connections can be made of the same length, optical logic circuits can be pipelined (even at the gate level). This means that, although the latency due to the interconnections might be quite long, the whole system can operate at the fastest data rate the gates can support.

The intrusion of optical interconnections into digital data processing may happen at different levels of hierarchy: Today, there are computer-to-computer interconnections. The next step will be chip-to-chip interconnections [1, 2, 6], where certain processing units on a chip are equipped with 'optical pins': i.e. with detectors for receiving signals and with light modulators or emitters for transmitting signals. Ultimately, these processing units may become smaller and smaller: Thus, we end up with optical gate-to-gate interconnections.

Current system architectures and design techniques for digital optical circuits interconnected at the gate-level [9 to 14] are conceptually different from classical computer design and VLSI: they are based on regularly interconnected two-dimensional arrays of devices, which are all identical with equal fan-in (= number of inputs to one gate), fan-out (= number of successive gates that one geate drives) and with only one type of logic function (for example NOR) across the array. The minimum fan-in and fan-out for a useful gate is two, since we need to be able to split and to combine signals. Homogeneously integrated arrays of infinitely cascadable gates with fan-in/fan-out of two are the mildest possible demand on the optical logic gates in order to be useful for digital optics. Devices that fulfill these minimum requirements have been demonstrated [16].

Large connectivity of the interconnections and large gate arrays pay off only if a significant fraction of the gates is busy at any one instant. Naturally parallel and pipelineable problems, such as switching in optical communications, profit directly from such an architecture. For more general problems the pipe must not break, which leads to a basic design rule: we need long, continuously connected circuits in order to prevent registering, which means preventing stop-and-go of the signal flow.

Symbolic substitution [9 to 13] is one approach to the design of pipelined, parallel and regularly interconnected circuits. The concept is general enough to design a complete processor. An algorithmic design technique that transforms arbitrary logic equations into a network of pipelined, parallel and regularly interconnected logic gates has been devised [14]. It is advantageous to incorporate into these designs global interconnection schemes, such as the perfect shuffle [3, 4], banyan networks [14] or the crossover [5]. These design techniques lead to programmable logic arrays similar to those known from electronics. But both of these techniques are in a beginning stage.

3. About Optical Switching Devices

Optical logic gates or optical switching devices are an essential ingredient for binary data processing: The gates perform logic, they open and close signal paths (routing of data) and, most importantly, they are responsible for the regeneration of the binary signal levels which get corrupted by system losses, noise and other imperfections. Amplifiers alone are insufficient since errors and noise accumulate during processing. In order to fully use the connectivity of optics it must be possible to fabricate two-dimensional arrays of switches. There are also practical considerations: the device should be manufacturable with reasonable yield, it should have wide tolerance margins, in temperature as well as in other operating conditions and it should run at room temperature.

Recently, many different optical switching devices have been demonstrated [15 to 21]. Some have sub-nanosecond switching times and dissipate per switching event an amount of energy comparable to electronics. They are based on bandfilling and/or excitonic nonlinearities near the band gap of III/V-semiconductors.

All-optical devices [17 to 21] often rely on Fabry-Perot resonators to enhance nonlinear changes of the refractive index of the material. There are optically bistable devices of this type as well as two-wavelength devices that can directly be used as optical logic gates [17, 18].

Opto-electronic devices [15, 16] based on multiple quantum well light modulators offer the possibility of interfacing directly GaAs-based electronic integrated circuits with parallel optical interconnects. Besides opto-electronic modulator devices bistable devices have also been demonstrated.

Opto-electronic devices may be preferable, at least today, to all-optical logic gates because part of the switching energy of the gate is provided electrically. Hence, the demand for laser energy is relaxed. In most studies about optical computing the optical power supply is neglected.

In the near future, integrated opto-electronic chips might be attractive for another reason: they consist of electronic processing units with optical terminals (detectors and modulators or lasers). As long as we cannot drive enough optical gates to build complex circuits, we still may interconnect electronic processing units optically. This way, we retain the established electronic processing power, however, we profit from the superior optical connectivity.

4. Energy and Speed

In general, it is necessary to make optical nonlinear devices very small (i.e. a few wavelengths in diameter) [18], in order to minimize the energy dissipation per switching event. Then, the switching speed is limited either by the carrier lifetime or by thermal transfer, that is the rate at which the dissipated heat can be removed from the chip. The speed, two-dimensional packing density and cooling rate set a general limit on optical data processing. For efficient cooling, reflection devices are preferable to transmission devices, which require either transparent cooling systems or can be cooled only from the edge of a chip. Let E_D be the dissipated energy per switching event and device area, I_C the power density that can be removed from a chip by the cooling system (without unacceptable temperature rise or thermal gradients) and α the packing constant (i.e. the ratio of active device area to chip area). Then the cycle time τ_C of the device is limited by:

$$\tau_C \leqq \frac{E_D}{I_C} \alpha .$$

Present cooling systems can remove on the order of 10 W/cm². More efficient cooling systems have been demonstrated [22]. The temperature rise (and the temperature dependence of the band gap of GaAs) may, however, be a major problem. GaAs-modulators typically [15] dissipate about 10^{-14} J/μm² per switching event. Hence, cycle times in the nanosecond range can be reached only at the expense of packing densities of $\alpha \approx 1\%$ only or less. Hence, reduction in energy dissipation is still an essential issue.

It is interesting to investigate how far away we are from the quantum limit: Assume N photons are absorbed for reliable switching with wavelength λ. Then, the minimum energy involved in a switching event is N times hc/λ. The minimum device area to which we can focus down to is estimated by $(\lambda/n)^2$, where n is the refractive index of the material. Hence, the single-mode switching energy density is estimated by:

$$E_D \approx N \frac{hcn^2}{\lambda^3}.$$

Comparing the material constant for GaAs $hcn^2/\lambda^3 \approx 4 \times 10^{-18}$ J/µm² with the experimental energy dissipation of about 10^{-14} J/µm² [15], we come up with some thousand absorbed photons per switching event and mode. Hence the photon shot noise is in the range of a few percent. Note by the way, that devices in the far infrared are more economic in terms of energy, since E_D scales with λ^{-3}: Working with $\lambda =$ $= 10.6$ µm (CO_2-Laser) [23] instead of $\lambda = 0.85$ µm (band gap of GaAs) would reduce the quantum limit for the single mode switching energy by a factor of nearly 2000.

5. Coupling Problems

Digital optical circuits will probably run on laser light: Optical switches often rely on resonant effects (band gap, Fabry-Perot or microresonators), which require temporally coherent light. In addition, devices of a few wavelengths in diameter require spatial coherence for efficient focusing. Upon an optical switch we have to combine at least two input-beams. If the device is not light-emitting by itself (such as an integrated phototransistor/LED combination), additionally a read-out or power-supply beam must be brought in. In a reflective device, the output has to be taken out from the same side as the inputs come in. To handle all those coherent, single mode signals without loosing much light or resolving power of the imaging systems poses formidable problems.

If all necessary signals would simply be combined on the device, interference-effects will modify the logic levels (1 + 1 becomes zero in case of destructive interference). To build complex optical circuits to interferometric accuracy seems impractical. Therefore, the different beams must be coupled into different modes of the device (different polarizations, at different locations, from different directions, at different times or with different wavelengths). Having more than one mode increases energy dissipation but is not avoidable.

Coupling into different orthogonal modes of a device allows lossless beam-combination, at least in principle. Systems for that purpose rely on polarizing beam splitters, pupil division [7], mirrors with spacevariant reflectivity [24] etc.

Another important system problem is undesireable feedback: the optical switches have to be isolated from switching events on successive devices, otherwise erroneous switching will occur. Antireflection coatings will help, but ultimately all the unused light must be dumped somewhere (not necessarily absorbed within the switch [17, 18]).

6. Characteristic Parameters of Optical Switches

In this section, the computational parameters of a switching device will be related to its optical parameters [7]. Fig. 1 illustrates these parameters for a typical input/output characteristic of a switching device. Here a non-inverting characteristic is shown. For an inverting device nothing significant is changed for the following analysis.

Computational parameters of the switching device are: fan-in — number of inputs sent into the device, fan-out — number of outputs available for successive devices, threshold — number of inputs necessary to switch the device.

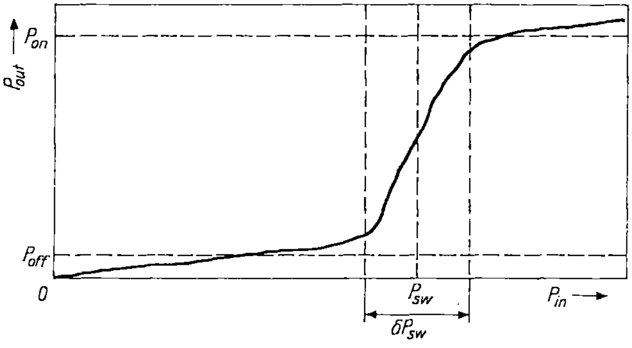

Fig. 1. Characteristic of a non-inverting switching device

Optical parameters of the switching device are: P_{on}, P_{off} — output power in the ON and OFF state, respectively, P_{sw} — input power at the threshold, δP_{sw} the necessary increment in input power for switch.

More compactly, the device is described by: $T_{dev} = P_{on}/P_{sw}$, the device transmission, $C_{dev} = P_{on}/P_{off}$, the device contrast, and $\sigma_{dev} = \delta P_{sw}/P_{sw}$, the relative switching increment.

Additionally, we need to take into account how the device is used. Properties of the optical imaging systems etc. are introduced by the following fudge factors; P_{bias} — an additional input beam to fix the operating point, T_{sys} — system transparency, takes into account all losses, σ_{sys} system accuracy, takes into account relative errors in laser power noise in the system, etc.

Hence, for reliable switching within a system the optical switch requires a minimum increment of input power given by $(\sigma_{dev} + \sigma_{sys})\, P_{sw}$.

7. The Gain-Condition

If a device switches, its output changes by $|P_{on} - P_{off}|$. This change has to be sufficient to drive fanout devices and suffers from the system losses. Hence, the input power of a successive device changes by

$$\frac{T_{sys}}{\text{fan-out}} |P_{on} - P_{off}| \approx T_{sys} P_{sw} \frac{T_{dev}(C_{dev} - 1)}{\text{fan-out } C_{dev}}.$$

Obviously, this input change must be bigger than the minimum switching increment. This yields the gain-condition:

$$\text{fan-out} \leqq \frac{T_{dev} T_{sys}}{\sigma_{dev} + \sigma_{sys}} \left[1 - \frac{1}{C_{dev}}\right].$$

A device-specific figure of merit, the gain-parameter of a switch, is given by

$$\beta = T_{dev} \left[1 - \frac{1}{C_{dev}}\right].$$

Later on the gain-condition will evaluated in Fig. 2a and b. Note, that absolute gain ($T_{dev} \geqq 1$) is not necessary, but it helps. High contrast and high accuracies can make up for missing gain. However, too stringent requirements defeat the idea of digital processing.

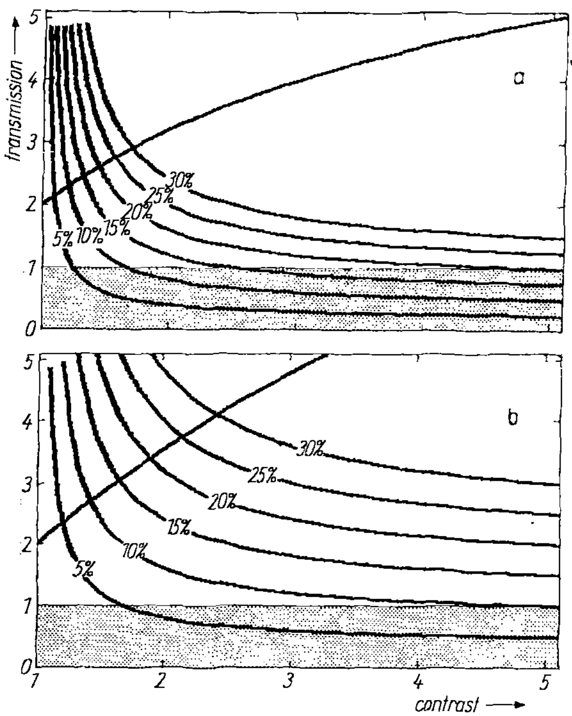

Fig. 2. Gain-condition and bias-condition for several system/device accuracies. As computational parameters a) fan-in/fan-out of two and a threshold of one were assumed, and b) fan-in/fan-out of four and a threshold of one were assumed. System transmission is 50%

8. The Bias-Condition

A second important condition relates optical parameters of a device with its fan-in: First we determine the power of the bias beam, which determines the operating point of the switch such that the switching power P_{sw} is reached, if the number of logically high inputs reaches threshold. If less than (threshold-1) inputs are in ON-state there is no switching:

$$P_{\text{bias}} + \frac{\text{fan-in}}{\text{fan-out}} T_{\text{sys}} P_{\text{off}} + (\text{threshold} - 1) \frac{T_{\text{sys}}}{\text{fan-out}} |P_{\text{on}} - P_{\text{off}}| \leqq P_{\text{sw}} - \frac{\delta P_{\text{sw}}}{2}$$

and if the threshold is reached there is enough power for switching:

$$P_{\text{bias}} + \frac{\text{fan-in}}{\text{fan-out}} T_{\text{sys}} P_{\text{off}} + \text{threshold} \frac{T_{\text{sys}}}{\text{fan-out}} |P_{\text{on}} - P_{\text{off}}| \geqq P_{\text{sw}} + \frac{\delta P_{\text{sw}}}{2}.$$

Both conditions together determine the optimal bias power, which must be non-negative. This yields the bias-condition:

$$\text{fan-in} \leqq \frac{\text{fan-out } C_{\text{dev}}}{T_{\text{sys}} T_{\text{dev}}} - \left(\text{threshold} - \frac{1}{2}\right)(C_{\text{dev}} - 1).$$

9. The Minimal Demands on an Optical Switching Device

In Fig. 2a and b the limits imposed by the gain- and the biasing condition are illustrated for different parameter choices. A minimal useful device requires a fan-in and fan-out of two, a threshold of one (it could be a two-input NOR-gate, for example), and it must make up for at least 50% system losses. Fig. 2a shows, for different accuracies $\sigma_{\text{dev}} + \sigma_{\text{sys}} = 5\%$ to 30%, the gain condition (hyperbolic curves) and the biasing condition (increasing curve). A switch is cascadable if it is located above the gain-condition-curve and below the biasing-condition-curve. The shaded rectangle corresponds to passive devices, without inherent gain ($T_{dev} \leqq 1$). If we allow system errors $\sigma \approx 10\%$, then a gain parameter of $\beta \approx 0.4$ must be provided. This can be done in terms of contrast or transmission. For $\sigma \approx 25\%$ a gain parameter $\beta \geqq 1$ is necessary, which is impossible for devices without inherent gain. Fig. 2b shows similar curves, however, fan-in and fan-out were chosen to be four. The tolerances become tighter in that case.

Unless we have devices with inherent gain, we have to live with low fan-in, low fan-out and with stringent systems accuracy requirements. The architecture must tolerate these device properties as well as the regularity of space-invariant optical interconnections. A device with inherent gain could be opto-electronic. One or several detectors which drive a light modulator or a light emitter. In this case, gain is provided electronically.

Acknowledgements

The authors wish to thank A. Huang, A. Lohmann, W. Stork, and a lot of other collegues in the field of digital optics for many stimulating and some useful discussions.

References

[1] R. KOSTUK, J. GOODMAN, and L. HESSELINK, Appl. Optics **24**, 2851 (1985).
[2] L. A. BERGMANN et. al., Opt. Engng. **25**, 1109 (1986).
[3] A. W. LOHMANN, Appl. Optics **25**, 1543 (1986).
[4] A. W. LOHMANN, W. STORK, and G. STUCKE, Appl. Optics **25**, 1530 (1986).
[5] J. JAHNS and M. MURDOCCA, Appl. Optics, to be published.
[6] K. H. BRENNER and F. SAUER, Appl. Optics, to be published.
[7] M. E. PRISE, N. STREIBL, and M. M. DOWNS, Opt. Quantum Electronics **20**, 49 (1988).
[8] G. E. LOHMANN and A. W. LOHMANN, Opt. Engng., to be published.
[9] A. HUANG, Proc. 10th IEEE Internat. Conf. Optical Computing, Cambridge (Massachusetts) 1983 (p. 13).
[10] A. HUANG, IEEE Proc. **72**, 780 (1984).
[11] K. H. BRENNER, A. HUANG, and N. STREIBL, Appl. Optics **25**, 3054 (1986).
[12] M. J. MURDOCCA, Appl. Optics **26**, 3365 (1986).
[13] K. H. BRENNER, Appl. Optics **27**, 1687 (1988).
[14] M. J. MURDOCCA, A. HUANG, J. JAHNS, and N. STREIBL, Appl. Optics **27**, 1651 (1988).
[15] D. A. B. MILLER, D. S. CHEMLA, T. C. DAMEN, T. H. WOOD, C. A. BURRUS, A. C. GOSSARD, and W. WIEGMANN, IEEE J. Quantum Electronics **21**, 1462 (1985).
[16] D. A. B. MILLER, J. E. HENRY, A. C. GOSSARD, and J. H. ENGLISH, Appl. Phys. Letters **49**, 821 (1986).
[17] J. L. JEWELL, Y. H. LEE, M. WARREN, H. M. GIBBS, N. PEYGHAMBARIAN, A. C. GOSSARD, and W. WIEGMANN, Appl. Phys. Letters **46**, 918 (1985).
[18] J. L. JEWELL, A. SCHERER, S. L. McCALL, A. C. GOSSARD, and J. H. ENGLISH, Appl. Phys. Letters **51**, 94 (1987).
[19] C. T. SEATON, S. D. SMITH, F. A. P. TOOLEY, M. E. PRISE, and M. R. TAGIZADEH, Appl. Phys. Letters **42**, 131 (1983).
[20] H. M. GIBBS, Optical Bistability, Controlling Light with Light, Academic Press, 1985.
[21] P. MANDEL, S. D. SMITH, and B. S. WHERRETT (Ed.), From Optical Bistability toward Optical Computing – The EJOB project, North Holland Publ. Co., Amsterdam 1987.
[22] D. B. TUCKERMAN, Thesis, Lawrence Livermore Laboratory, 1984.
[23] L. C. WEST, Thesis, Lawrence Livermore Laboratory, 1985.
[24] M. E. PRISE, M. DOWNS, S. WALKER, and N. STREIBL, Proc. Conf. Optical Bistability, Aussois (France) 1988.

(Received August 25, 1988)