

---

# Impact of Technological Advances and Architectural Insights on the Design of Optical Computers

A. Huang

*Phil. Trans. R. Soc. Lond. A* 1984 **313**, 205-211  
doi: 10.1098/rsta.1984.0097

---

## Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

---

To subscribe to *Phil. Trans. R. Soc. Lond. A* go to: <http://rsta.royalsocietypublishing.org/subscriptions>

---

## Impact of technological advances and architectural insights on the design of optical computers

BY A. HUANG

*AT&T Bell Laboratories, Crawford Corner Road, Holmdel, New Jersey 07733, U.S.A.*

Communication problems such as interconnection bandwidth, clock skew, and connectivity are restricting computational throughput. Bandwidth and clock skew problems limit the speed and add to the design complexity of a processor. Constrained connectivity forces much of the speed of a processor to be used to compensate for the limited number of interconnections.

Philosophically, the large bandwidth, innate parallelism and non-interfering propagation of optics offer mechanisms for overcoming these communication problems. The difficulty in exploiting these capabilities has been the absence of suitable optical logic and memory devices. Advances in optical nonlinearities offer the possibility of cascaded optical logic gates that are competitive with electronics. Advances in computer architecture can be used to simplify the optical memory requirements and utilize the large bandwidth, parallel, non-interfering communications of optics.

People have been interested in the possibility of an optical digital computer for some time. Previous efforts were limited by the absence of suitable logic and memory devices as well as the lack of a clear understanding of what optics has to offer and what the computer community needs. The technological spin-offs associated with the development of optical communications now offer the possibility of optical logic with speeds and power comparable with conventional electronic logic. What is missing is a better understanding of the capabilities of optics and the problems facing future computer systems.

### PHYSICAL CONSTRAINTS ON COMPUTATIONAL THROUGHPUT

One problem confronting future computer systems is interconnection bandwidth. As system cycle times and pulse widths shrink, the bandwidth needed to preserve the rising and falling edges of these signals increases. This forces the need for bulky, expensive, terminated coaxial interconnections.

Another problem is clock skew. This problem occurs when signals from different parts of a circuit arrive at a gate at different times. This skew of the inputs can cause a gate to generate an erroneous output. The desire for shorter cycle times limits the amount of clock skew that can be tolerated. This in turn limits the maximum difference in interconnection length. A processor such as the Cray-1 has a maximum interconnection length of six inches (*ca.* 15 cm). All interconnections less than this length must be padded with gate delays to make the propagation time equivalent to that of the maximum interconnection length. This restriction complicates both the physical and electrical designs of a processor and hints at some of the difficulties associated with the next generation of processors.

The fear of clock skew precludes the use of logic in a pulsed mode. The accepted approach when using digital logic is to wait for the inputs to settle before using the output of a gate. This input settling time is dependent on the amount of time it takes to fully charge the connection. In most circuits, the  $RC$  time constant-dominated settling time is longer than the transistor switching time. This makes it difficult to utilize ultrafast logic gates.

The difficulties associated with settling time are not solved by very large scale integration. As the length of a wire shrinks by a factor of  $\alpha$  and the cross-sectional area of the wire is reduced by a factor of  $\alpha^2$ , the capacitance of the wire decreases by a factor of  $\alpha$  while the resistance increases by the same amount. The  $RC$  time constant remains the same and thus the input charging time remains unaltered, independent of scaling (Mead & Conway 1980). Given the parameters of very large scale integration it is estimated that the signals will be communicated at approximately 0.5% the speed of light (Wilkes 1983).

#### THE 'VON NEUMANN BOTTLENECK': THE USE OF TIME TO REDUCE INTERCONNECTIONS

Conventional computers suffer from a 'bottleneck' that is caused by the limited number of interconnections that can be supported in a practical manner by electronics. This problem is referred to as the 'Von Neumann bottleneck' and involves the performance limitations imposed by the sequential and address-oriented communications between the *central processing unit* and the *memory* in a conventional computer (Backus 1982).

The source of this bottleneck can be found by examining the 'classical finite state machine', an ancestor of modern computers. This processor, shown in figure 1, consists of storage elements,

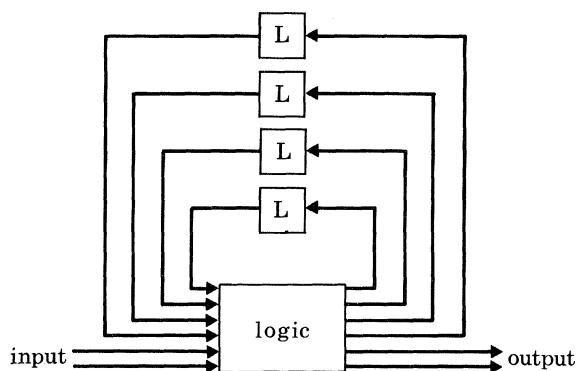


FIGURE 1. A 'classical finite state machine' does not suffer from the Von Neumann bottleneck since it can update all of its memory in parallel without the need for addresses.

a combinatoric logic unit, inputs, outputs and various interconnections. What is unusual about this processor is that it does not suffer from a Von Neumann bottleneck. All storage elements are updated in parallel without the need for addresses. The bottleneck emerges when more storage variables are added. It becomes impractical to support  $N$  interconnections between the storage and logic units with wires. As a result, the classical finite state machine is modified to the structure shown in figure 2, in which a binary encoding scheme is used to reduce the interconnections from the output of the logic unit to the memory and a common return line

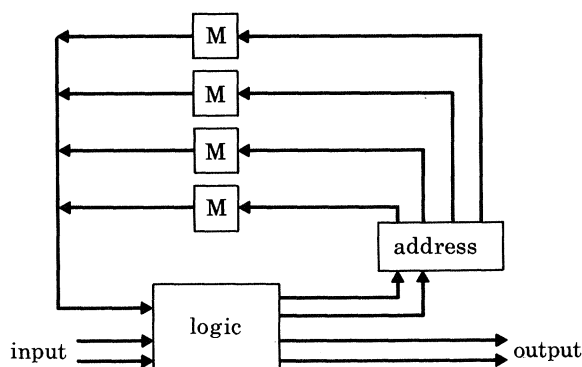


FIGURE 2. A 'modified finite state machine' suffers from the Von Neumann bottleneck since it can only update one memory element at a time and it needs an address to do so.

is used to reduce the interconnections between the memory elements and the input of the logic unit. This reduces the number of interconnections, but it also seriously degrades the performance since the 'modified finite state machine' can only address one storage element at a time and an address is needed to specify which element. This inability to support  $N$  interconnections in parallel is the origin of the Von Neumann bottleneck.

#### OTHER BOTTLENECKS THAT USE TIME TO REDUCE INTERCONNECTIONS

Computers suffer from other constrictions similar to the Von Neumann bottleneck. For ease of design and construction a processor is partitioned into modules. Inevitably, each module must communicate with the other modules. The impracticality of fully interconnecting  $M$  modules with  $M \times (M - 1)$  bus-wire interconnections leads to the use of a broadcast bus structure that uses time to reduce the number of interconnections. This results in sequential, address-oriented communications identical to those of the Von Neumann bottleneck.

A similar bottleneck occurs at an even lower level. Ideally, it would be nice if each bit of a memory chip could be independently read or written. To accomplish this a  $P \times 1$  bit memory chip would have to have  $P$  input and  $P$  output lines. Since this is impractical, a binary encoding scheme is used to reduce the  $P$  inputs to  $\log_2 P$ . This approach is similar to that used in the modified finite state machine and results in another sequential, address-oriented communications bottleneck.

These communication bottlenecks at the architectural, bus and chip level all stem from the use of time multiplexing to compensate for an inability to communicate  $N$  channels of information in parallel. These bottlenecks, along with the previously mentioned bandwidth and clock skew problems, are interrelated. Increasing the switching speed might ease the connectivity problem, but it also aggravates the bandwidth and clock skew problems.

#### HOW CAN OPTICS HELP?

Optics is capable of communicating many high bandwidth channels in parallel. Lenses, prisms and mirrors can convey images consisting of millions of resolvable spots. Each spot is capable of supporting a very large bandwidth channel. Optics also has the benefit of

non-interfering propagation. Optical beams can cross without interaction. These attributes have been exploited in analog optical signal processing, but they have yet to be applied to digital processing. The main reason for this has been absence of suitable optical logic and memory devices.

#### OPTICAL LOGIC

Many optical logic gates have been demonstrated and discussed in the literature. The main problem is that these gates are not cascadable (Basov 1972). The inputs of these gates are represented by a phenomenon such as phase while the output is expressed in another phenomenon such as intensity. As a result, the output of one gate cannot be used as the input of another gate.

The prospects for a cascadable optical logic have recently changed (Abraham *et al.* 1983). Optical nonlinearities that have previously only been observed at high power levels have now been demonstrated at switching energy levels comparable to transistors (Smith 1981). Nonlinearities requiring 10 s of watts per bit but reacting in 100 s of femtoseconds, as well as nonlinearities requiring  $10^{-8}$  W bit<sup>-1</sup> and reacting in 10 s of nanoseconds, have been observed (Gibbs *et al.* 1982; Miller 1982, 1983). Primitive cascadable AND, OR, NOR and NAND gates have also been demonstrated (Abraham *et al.* 1983).

#### OPTICAL MEMORY DEVICES

The use of optics in a computer has also been hindered by the lack of a suitable optical memory. Several optical memories were developed but they were slow, awkward and expensive. In retrospect, the optical memories that were developed were for a modified finite state machine rather than a classical finite state machine. This seemingly minor difference sacrificed the parallelism of optics and forced optical memories to incorporate addressing mechanisms consisting of beam deflectors, page composers and detector arrays. Another important distinction between these architectures is the type of storage required. The storage elements of a modified finite state machine must be capable of preserving information indefinitely since the elements are addressed in random order. The search for an optical material capable of storing information in this manner has proven quite difficult. It would have been considerably simpler for optics to implement the memory as required by a classical finite state machine, since the storage elements need only preserve their information for one cycle. This architecture would also have used the parallel communications capabilities of optics and avoided the need for an addressing mechanism.

A direct descendant of the classical finite state machine is a 'parallel pipelined processor' as shown in figure 3. The alternating levels of latches and logic of the parallel pipelined processor can be viewed as an unravelled version of the classical finite state machine. The latches of the parallel pipelined processor provide the necessary storage, do not require an addressing mechanism, and are simpler to implement for optics since they need only hold information for one cycle.

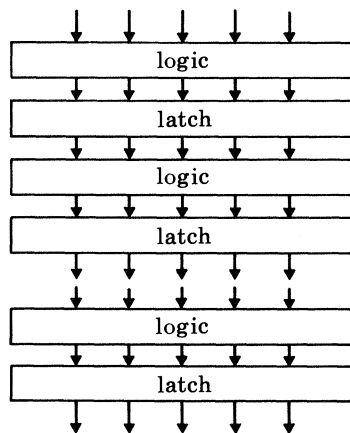


FIGURE 3. A 'parallel pipelined processor' with its alternating stages of latches and logic can be viewed as an unravelled version of a classical finite state machine.

#### BENEFITS OF A PIPELINED ARCHITECTURE

Power considerations have been used to dismiss optical logic and play down the potential femtosecond switching rates of optics (Basov 1972; Keyes 1975). A pipelined architecture can be used to reduce this problem. Only a small fraction of a conventional central processor is being used during any given cycle. A considerable amount of power can be saved by turning off the unused circuits. One way of accomplishing this is by casting the processor into a pipelined architecture and only partly filling the pipeline.

The memory of modern computers also consumes considerable amounts of power. Given an ultrafast logic it becomes practical to consider the use of the non-dissipative propagation delays to provide the latching as required by a pipelined architecture. This might seem a bit unusual, but it should be remembered that mercury delay lines were used to provide storage in early computers.

A pipelined architecture can also be used to structure the interconnections (Huang 1984). This simplifies the task for optics and also opens the possibility of using femtosecond optical nonlinearities. The difficulties associated with trying to communicate femtosecond signals without clock skew has discouraged the investigation of logic devices based on these phenomena. While it is difficult to route and fabricate equidistant waveguides, it should be remembered that mirrors and lenses can provide equidistant optical paths by using free-space propagation. This oversight illustrates some of the difficulties in extrapolating between technologies. One of the hidden assumptions was that since electronics must use wires to communicate, optics must use waveguides. A second assumption was that a digital circuit always involves an awkward mix of interconnection lengths.

#### OPTICS COMPARED WITH ELECTRONICS

The question as to which technology switches the fastest is only part of the overall problem. An analysis of the switching speed of various technologies must be coupled with a measure of the communications ability of that technology. The previous discussion pertaining to the Von Neumann bottleneck demonstrates how an inability to communicate can quickly compromise

any advantage a technology might have in terms of speed. Optics has a definite advantage over electronics in terms of both the bandwidth and number of free propagating channels that can be supported per given volume.

At first glance, the relative merits of electronics and optics seem to follow from the basic characteristics of electrons and photons. Electrons affect each other even at a distance. As a result, it is easy for one electrical signal to affect itself as well as another signal. This makes it easy to perform switching, but this interaction complicates the task of communication. Interaction in the form of inductance and capacitance limits the propagation speed and bandwidth of signals. Interaction in the form of electrical or magnetic coupling threatens the integrity of a signal. Photons behave quite differently. It is very difficult to get two photons to interact. This makes it difficult for optics to perform switching, but this lack of interaction simplifies the task of communication. The symmetry of this conjecture is very appealing, but advances in optical nonlinearities show that optics can be used to perform switching at energies comparable to electronics. This puts electronics with its communication difficulties at a disadvantage.

#### ECONOMICS

Economics has always played a part in determining the viability of optical digital computation. It is important to consider how optics can compete against the billions of dollars already invested in electronics. These concerns have prompted the search for problems that would need and could pay for the capabilities of optics. Trying to sell a problem, a solution and a new technology at the same time is difficult. This situation is being changed by the optical communications revolution. Long haul communications are being dramatically changed by light-wave technology. The large data rates involved will force similar changes in the communications between processors. These changes will encourage changes in the communications between modules within a processor and will eventually force the communications between chips and even devices on the same chip to rely on optics. Optical local area networks and optical computer buses are already being developed. Ways of using optics to increase the communications on and off of chips are being vigorously pursued. The insatiable demand for increased data rates makes this trend inevitable. This evolutionary path is self-perpetuating. Optics is generating problems that only optics can solve. This trend will pay for and generate a new optics technology.

#### SUMMARY

Current computers suffer from several communication problems that limit their computational throughput, such as bandwidth, clock skew and the Von Neuman bottleneck. The bandwidth and clock skew problems limit the speed and add to the design complexity of a processor. The Von Neumann bottleneck forces much of the speed of a processor to be used to compensate for the limited number of interconnections. Philosophically, the large bandwidth, innate parallelsim and non-interfering propagation of optics offer mechanisms for overcoming these communication problems.

Historically, the development of optical digital processors has suffered from the lack of suitable logical and memory devices. Recent advances in optical bistability offer the possibility

of cascadable optical logic gates with speeds and power consumption competitive with electronic gates. A parallel pipelined architecture can be shown to simplify the optical memory requirements; utilize the large bandwidth, parallel, non-interfering communications of optics; and solve some of the expected problems associated with using optical technology that operates on a femtosecond timescale.

## REFERENCES

- Abraham E., Seaton, C. T. & Smith, S. D. 1983 *Scient. Am.* **248**, 85–93.  
 Backus, J. 1982 *IEEE Spectrum*, **18**(8), 22–37.  
 Basov, N. G. 1972 In *Laser handbook* (ed. F. T. Arecchi & E. O. Schulz-Dubois), pp. 1649–1693. Amsterdam: North Holland.  
 Gibbs, H., Tarng, S. S., Jewell, J. L., Weinberger, D. A., Tai, K., Gossard, A. C., McCall, S. L., Passner, A. & Wiegmann, W. 1982 *Appl. Phys. Lett.* **41**, 221–222.  
 Huang, A. 1984 *Proc. Inst. elect. Electron. Engrs* **72**, 780–786.  
 Keyes, R. 1975 *Proc. Inst. elect. Electron. Engrs*. **63**, 740–767.  
 Mead, C. & Conway, L. 1980 *Introduction to VLSI systems*. Reading, Massachusetts: Addison-Wesley.  
 Miller, D. A. B. 1982 *Laser Focus* **18** (4), 79–84.  
 Miller, D. A. B. 1983 *Laser Focus* **18** (7), 61–68.  
 Smith, P. W. & Tomlinson, W. J. 1981 *IEEE Spectrum* **19** (6), 26–33.  
 Wilkes, M. V. 1983 *Conf. Proc. 10th Ann Int. Symp. on Computer Architecture*, IEEE cat. no. 83CH1889-5, pp. 2–4