1

LEVELS OF ABSTRACTION IN COMPUTING SYSTEMS AND OPTICAL INTERCONNECTION TECHNOLOGY

Haldun M. Ozaktas

Bilkent University
Department of Electrical Engineering
Turkey

ABSTRACT

The design of a computing machine takes place at several levels of abstraction ranging from materials and device engineering to system architecture to high-level software. This system of levels of abstraction enables the design problem to be broken down into manageable subproblems, much as in a procedural programming language. On the other hand, it makes difficult the introduction of novel concepts and technologies such as optoelectronic device planes ("smart pixels"), which do not readily fit in the existing scheme of things. We try to develop an understanding of this system of levels of abstraction, why and how it resists the introduction of optical technology, and how one can modify it so as to successfully house optical technology. We argue that in the near future, optoelectronic technology can be successfully introduced if: (i) changing technology or applications create a significant bottleneck in the existing system of levels of abstraction that can be removed by the introduction of optical technology (e.g. interconnections, memory access); (ii) special purpose applications involving very few levels of abstraction can be identified (e.g. sensing, image processing); (iii) it is possible to modify a few levels of abstraction above the level that optical technology is introduced, so that the optical technology is smoothly "grafted" to the existing system of levels of abstraction (e.g. modifying communications schemes or standards so as to match the capabilities of optical switching systems, employing parallel architectures to match the parallel flow of information generated by optical subsystems).

1   INTRODUCTION

The integration of larger numbers of primitive computing elements (switches, transistors, gates, processors, etc.) to produce computers of greater processing power requires the use of interconnections with greater length/width ratios.[1]

As the length of an interconnection is increased, the time it takes for a signal to propagate to the other end also increases, at least as much as dictated by the speed of light. While this limitation holds for all types of interconnections, normally conducting electrical interconnections have much more severe limitations. The signal delay is a quadratic function of the length/width ratio beyond a certain length/width ratio, since the line becomes too lossy to allow pulse propagation. The energy per transmitted bit also increases with line length, even when repeaters are used [8,11].

For these and other reasons (e.g., the possibility of non-planar interconnections, voltage isolation, very little or no frequency dependent crosstalk and distortion, no impedance matching problems even with multiple taps, etc.) that have been extensively discussed elsewhere, it has been suggested to use optical interconnections for implementing the longer connections in computing systems, especially when an electrical line to be used instead would have a high length/width ratio. (See [1–6,8,9,11–14] and the references therein.)

From a fundamental perspective, the advantages of optical interconnections in comparison to conducting interconnections is almost obvious to anyone with a basic understanding of the physics involved. Many analytical and quantitative studies, as well as some technology demonstrations also confirm the advantages to be gained by employing optical interconnections. Despite this, the penetration of optical technology into mainstream computing systems has been disappointingly slow. In this essay, we would like to take the opportunity to explore the reasons for this in a qualitative manner. Given the nature of the problem we try to address and the subjective and descriptive style of this essay, we would like to acknowledge its limitations, incompleteness, and need for refinement at the outset.

---

[1] This can be avoided by resorting to architectures with local connections only, but for problems which intrinsically require global flow of information this merely amounts to breaking down the necessary long distance communication paths into a large number of short hops, which is not necessarily optimal [13].

2 LEVELS OF ABSTRACTION IN PROBLEM SOLVING

A lot of attention has been paid to determining at what level of the interconnection hierarchy optics should be employed (board to board? chip to chip? etc.). On the other hand, it seems that little attention has been paid to the issue of determining at what level of abstraction of the computational process optics should be introduced. (It will be evident that the two issues are not unrelated, the levels of abstraction to some degree corresponding to the levels in the packaging hierarchy.)

To understand what this means, let us reflect on how one usually accomplishes a task by using a computer. Say that we wish to modify a high resolution image in some way so that it is more pleasing to the eye. It might be possible to come up with certain operations involving convolutions, matrix operations etc. that enable us to do this. These mathematical operations will have to be broken down into repetitive or recursive sequences of more elementary operations. In doing this, we will have developed an algorithm for solving the problem. The algorithm will be written down in the form of a high-level programming language, which will be translated into assembly language, which will be run by a microprocessor, which is essentially a high-level logic system, which is made up of lower-level logic functions (such as shift, add, etc.), which are made of gates, which are made of transistors.

Given a certain number of transistors or logic gates, there is no reason to think that this way of doing image enhancement is optimal, but at least it is possible. Realization of the techniques associated with each level of abstraction can be posed as self-contained problems which can be solved by specialists, with some care being necessary to ensure successful interface to the levels immediately above and below. Some degree of optimization within each level is usually performed, but it is often not possible to optimize over several levels. No central committee has ever decided on what these levels of abstraction are either; they are the outcome of historical developments.

Device physicists try to minimize the switching time and energy, and computer scientists try to minimize the consecutive number of steps required for the completion of a task. This suggests that the combined effort of both camps will result in an optimal machine, but closer examination reveals that things are not so simple. For instance, perhaps it is the case that for larger systems, performance saturates with increasing device speed, and devices beyond a certain speed offer no further increase in performance.

The controversy over the relative virtues of global and local computation cannot be resolved unless optimization over several levels of abstraction are performed (see [13] and references therein). Globally connected systems allow fewer steps of computation but result in longer duration per step, whereas locally connected systems require a larger number of steps with shorter duration per step; these considerations being closely related to the choice of algorithm, architecture, and interconnection media. To find the optimum degree of globality or locality, one must optimize jointly over possible algorithms and physical realizations of the machine.

The difficulty of introducing optical technology despite its clear fundamental physical advantages can to some extent be explained in the light of the above discussion. If we had a theory of computing which allowed joint optimization over all levels of abstraction, we could throw in the possibility of optical interconnections and switching into the parameter space. Then, given a computing task, we would perform the optimization, which would not only clearly indicate whether and when we should use optics, but also the architectures and paradigms that must be used. Since we cannot do this, we instead try to show that, say, a globally connected interconnection network is faster if implemented optically. But what if a locally connected network, which can be implemented electrically, allows the same task to be done in overall less time by running a different algorithm? (We have argued that this is not the case in [13], but not definitively.)

3   ALTERNATIVE SYSTEMS OF LEVELS OF ABSTRACTION FOR GENERAL PURPOSE COMPUTING?

It is clear that a very fast, large, and low switching energy array of optical switches or "smart pixels" has tremendous computing potential. However, it is too difficult a task to start from this array and arrive at a general purpose system in a single leap. If we are interested in designing a general purpose computer, we must guide our efforts by some system of levels of abstraction. It is first necessary to show how certain elementary functional units (in the abstract sense) can be formed, and then how these can form higher-level units and so on, until we arrive at some kind of high-level "programming language" enabling the problem description to be formulated. (In most cases, the burden of providing a higher-level platform, which must rest on intermediate-level

platforms, will belong to whoever provides the computer. People will certainly be reluctant if they are presented with an array of optoelectronic devices, no matter how fast or large, if we cannot show them where to plug the keyboard and monitor, and where to buy a C Language compiler. Without any registers, accumulators, microprocessors, assembly or C Languages, no user will want to program or configure their systems at such a low level.)

We could try to come forward with a new system of levels of abstraction complete with the techniques necessary for realization of each level, and then build machines including optical components based on this system. Taking the array of optical switches as our starting point, and without being biased by the mainstream system of abstraction, we may try to work our way up to the level of problem description.

Some alternative systems of abstraction do already exist, such as cellular automata, connectionist systems, and most significantly parallel computing. There is some reason to think that these might house optical technology better, but unfortunately these "paradigms" (which differ from the mainstream in varying degrees) are not that well developed. For instance, the techniques for only the lowest levels of abstraction are developed for cellular automata; nobody has a high-level programming language which they can compile into some kind of "assembly language" which will run on some kind of cellular automata hardware (which consists of several levels of abstraction down to the level of a single cell). The state of development of techniques for doing things with cellular automata is comparable to that of low-level logic in the mainstream system, such as shift registers etc.[2]

In conclusion, it seems difficult to come forward with a general purpose optically interconnected computer based on such novel paradigms; the development of the mainstream system of abstraction having spanned at least a century.

---

[2] It has been shown how to simulate conventional logic operations in cellular automata, so that one can in principle do anything with a cellular automaton that one can do with conventional logic. However, this is a meaningless approach if the cellular automata is implemented using logic gates in the first place, or simulated on a workstation. But things may change if cellular automata are implemented by virtue of some atomic scale physical phenomenon.

4 GENERAL PURPOSE SYSTEMS VERSUS SPECIAL PURPOSE SYSTEMS

Unlike general purpose systems which can be programmed to do any task with reasonable efficiency,[3] special purpose systems are "hardwired" and can do only certain prescribed tasks. Broadly speaking, the more general purpose a computer is—that is, the greater the range of tasks it can do with reasonable efficiency—the greater the number of levels of abstraction it has. The more special purpose a computer is, the fewer (for instance, the abacus does not have many). Thus, special purpose systems provide a better opportunity for new technologies (such as optics) and underdeveloped paradigms (such as cellular automata).[4]

Midway between the extremes of special purpose and general purpose systems we can identify a class of systems which we may refer to as "quasi-general purpose" systems or "coprocessors." Such systems can perform a certain class of operations of general utility, such as math coprocessors or digital signal processing chips. Of course, the full picture is that there is a continuum of systems of varying degree of "general purposeness" between the two extremes of special purpose and general purpose systems.

---

[3] It does not take much to be able to do any task, if one allows for gross inefficiency.

[4] It is quite conceivable for a limited number of, say, image processing researchers to start from a description of the capabilities of an array of optical devices and devise algorithms and methods for performing tasks they are interested in. (Many researchers were interested when the systolic computation paradigm was introduced for VLSI systems, unveiling a new class of solvable but unsolved problems.) The key issue seems to be that it should be possible for a single group or working unit to be able to obtain fruitful results by themselves, since this will give them the incentive to attack the problem. On the other hand, the effort towards the general purpose system would require a much bigger effort, requiring strategic commitment by a larger institution.

Researchers and engineers make careers out of solving the problems associated with a certain level of abstraction in the mainstream system. They will not be willing to change their focus easily, since within the present system, the people working at the lower level are providing them with the technology to realize their stuff, and the people at the higher level want the stuff to realize whatever they are doing at their own level. No one will benefit from change unless everybody changes at once. This is a particularly severe kind of "chicken-and-egg problem," since it will not by itself change for the better once given a sufficient but small initial momentum. On the other hand, there are always people willing to work on special purpose systems, which due to their limited number of levels of abstraction, can be handled independently by a single person or group. Thus, successful exposition of the capabilities of optical technologies to the image processing and computer science communities may be rewarding.

An issue which perturbs these considerations is the fact that in areas of academia where theoretical achievements are valued, the interests of researchers may be independent of whether they can interface with upper or lower levels of abstraction.

The "programming" of a general purpose system can take place at various levels. A microprocessor is a custom designed chip which can be programmed at a fairly high level (assembly language). On the other hand, a system to do the same task can be programmed at the much lower hardware level, for instance, by customizing a gate array. Both the microprocessor and the gate array can be viewed as general purpose systems which can be programmed to perform special purposes; the difference is in the nature of the programming and the level at which it takes place.

Under the light of what has been discussed until now, it is no wonder most successful optical systems to date have been special purpose systems. Such systems can be designed to perform a certain task standing alone, or they might be designed as a self-complete component of a larger computing machine. There is nothing complicated with the former. As a very simple example, an array of optical devices might be used as an image amplifier in a medical imaging system. The intricacies involved in the latter case will be discussed later.

Optoelectronic systems such as memory with parallel access, state machines, matrix processors, neural networks, etc. may represent more realistic challenges as compared to general purpose systems, for short term development and validation of optoelectronic technologies. These systems do not involve too many levels of abstraction (which makes their conception possible), often involve regular patterns of information flow (which leads to simple physical architectures), and usually result in an interconnection bottlenecked system when implemented with purely electronic technologies. The major challenge with such a system is to either successfully interface it as a subsystem of a larger (possibly general purpose) system in a way that benefits from its high performance, or to find a special purpose application where it can directly exhibit its high performance.

An interesting case is that of subsystems with few inputs and outputs, which we will discuss in the remainder of this section.[5] In general one would expect a function with a small number of inputs and outputs to be implemented with a small number of components. However, in certain special cases it might be possible to come up with an efficient implementation involving redundant replication, outer product generation, etc. of the data, followed by some fairly simple or regular processing in parallel, followed by reduction to the desired answer. Such a subsystem would more easily fit into existing computing systems without requiring major architectural changes at the higher levels, since the parallelism of optics is exploited in an entirely transparent way, and does not lead to any interface problems.

---

[5] This case was pointed out by David A. B. Miller.

One example is the optical method of correlation where the input function is replicated in shifted versions, multiplied by a replicated mask, and then integrated. Digital optical implementation should be possible by performing the spreads and integrations in $\sim \log_2 N$ stages. Searching a large database is another example. Assume that we have a certain number of subject terms and we wish to retrieve the entries containing these subject terms. The input and output are small, but the search must take place through a large space. Yet another example is matrix-vector multiplication with a fixed matrix. It should be possible to increase the number of individual examples, but more important is to find a general class of such problems which have some central significance.

When an optical solution of this kind is found, we must immediately inquire whether more efficient electronic implementations exist, since the small number of inputs and outputs suggest that it may be possible to implement the desired system with a small number of components. For instance, systolic convolution or correlation on a linear array can be performed in the order of N time. If the input to the subsystem is arriving in serial manner, it will take this long to read it in anyway so that the optical method will not present any advantages. If, however, the input vector is available in parallel, there is a chance that the optical method might offer some advantages such as lower cost or greater speed.

5   INTRODUCING OPTICS INTO GENERAL PURPOSE SYSTEMS

Although we have seen that special purpose systems provide a conceptually simpler opportunity for optical technologies, we wish to explore how optics may be introduced into general purpose computing systems as well. Since we have seen that it is very difficult to come forward with an all-together novel system of levels of abstraction which would house optical technology in an efficient way, it is clear that general purpose computers will be mostly based on the mainstream system of levels of abstraction (which we might be able to modify to a limited degree).

First, let us consider modifying only the least abstract level of problem solving, the level of physical devices, wires, etc. In this approach, we start with the mainstream architectural and packaging paradigm and see whether it is possible to make a "better" machine by using optical components (interconnections and/or switches) instead of some of the electrical ones. Examples of this approach might be the introduction of optical backplanes or chip-to-chip

modules instead of their electrical counterparts, while leaving the architectural conception and logical structure of the machine intact. This would change the job of the device physicist[6] and the person who designs the physical packaging, but would not affect people working at higher levels of abstraction, including those contemplating the logical and systemic architecture of the machine, as well as those providing the software.

This approach is appealing in that we do not have to worry about the development of new architectural concepts. However, there is no reason why the existing concepts should be particularly congenial to optical technology. In fact, they have historically developed to benefit from the strengths and accommodate the weaknesses of electrical technology, which are in some senses complementary to those of optics, so that this approach may not bring out the best of optical components. (VLSI architectures which try to minimize the length and number of chip to chip interconnections provide a good example.) Nevertheless, this may still be a valid and promising approach because it seems that replacing the longer wires with optical links does indeed result in a net advantage, even in existing systems.

If instead of the above simple approach, we wish to modify higher and higher levels of abstraction with the hope of better utilizing the particular optical technology at hand, we must face and overcome certain difficulties. For instance, we may attempt to replace a complete electronic combinatorial or sequential logic unit with an optical one which provides the same functionality, but in a "better" way. The interior structure and levels of abstraction of the optical unit may be entirely different, but it must interface with the system of levels of abstraction of the machine in which it is embedded at a certain level.

At relatively high levels of abstraction we might contemplate an optical microprocessor or digital signal processing coprocessor. At yet higher levels the physical and logical architecture of the machine will be altered significantly to suit the strengths of the optical technology. For instance, we might contemplate an optically interconnected parallel random access machine where the processor locations and algorithms are designed so as to match precisely the type of connection patterns that can be efficiently provided by optics.

Modifying the system at higher and higher levels of abstraction so as to better suit the optical technology becomes an increasingly difficult task as we move upwards because of the need to maintain continuity between the different lev-

---

[6] More precisely, it would create jobs for some device physicists while eliminating jobs for others.

els. If the central processing unit of a machine works on 32 bit wide words, its replacement must also work with 32 bit words. (It must be "plug compatible.") As another example, if we are to replace the existing electronic memory with an optical one, the input-output characteristics of the new memory must match those of what is being replaced. Notice that this requirement may sometimes resist improvements. A new optical memory, which provides much faster parallel access, may offer no system improvement, since the system in which it is embedded may not be able to utilize it. This makes it difficult to justify the optical technology, since the potential increase in performance offered by the optics cannot be utilized in this case, while its usually greater price will have to be paid. (On the other hand, perhaps an optical processor can be used to perform, say, some kind of parallel search on the data read from the optical memory, a feat which would be very expensive or slow with an electronic memory. The question now is whether the next higher level of the system can beneficially use the results of this fast parallel search. The answer would probably be yes if the search query as well as the result consist of small amounts of data.)

If we cannot succeed in getting a successful interface at one level, we might have to move up to a higher level and try our chance at that level. By modifying this higher level (which may or may not involve optical components), we might be able to exploit the higher parallelism or bandwidth offered by optics. If not, we might have to move another level up, until the intrinsic advantages of the optical technology seep through to the surface and translate directly into a user-level performance advantage (such as getting the job done in less time).

This discussion should also clarify what is meant by doing something "better." Doing some intermediate-level operation cheaper, faster, larger, etc. by introducing certain modifications at that level do not automatically result in user-level improvements. It may be necessary to make further modifications at higher levels, until the fastness, cheapness, etc. can seep through to the user level (which is the highest level).

Modifying the system at a certain level of abstraction might mean introducing an optical subsystem (such as an optical logic unit) into the machine, but this need not be the case. Remember that replacing all of the electronic switches and wires with optical ones does not alter the system of levels of abstraction at all, although we now have a computer consisting entirely of optical components. On the other hand, we may modify the architecture of the machine drastically, without introducing any optics at all. Despite the fact that the interconnection and packaging hierarchy often mirrors the levels of abstraction, the two concepts are distinct and must not be confused.

We now discuss a few examples to make the content of the last few paragraphs more concrete.

5.1   High-bandwidth "transparent" photonic switching

Research in guided-wave wideband switching networks has resulted in rather impressive switches whose various strengths and weaknesses are not exactly matched to the requirements of existing multiplexed switching networks, so that it seems they may find less application than originally hoped for. The weakness of these switches is that they have a limited number of spatial channels. Their strength is that they can route very high-bandwidth signals transparently. Efficient use of this bandwidth cannot be made if bitwise multiplexing is employed, since these systems cannot switch at a rate as high as their transmission bandwidth. However, if we make the higher-level modification of employing large-size block multiplexing instead of bitwise multiplexing, we can walk around this disadvantage. Now we must face the issue of whether the use of large-size block multiplexing is compatible with the next higher level of abstraction (which might be that of communications protocols and transmission standards). If not, we may try to push forward by suggesting modifications to the protocols and standards. If we do not arrive at a clear advantage within a few levels, we might have to give up.

5.2   Two-dimensional digital optical image processing

Optical technology will probably make it possible to construct image processing subsystems which can perform two-dimensional signal processing operations (such as the discrete Fourier transform (DFT), convolutions, etc.) in parallel at a very fast rate. Given the fact that digital electronic hardware is extremely strained to perform such operations of even moderate complexity, it initially seems that digital optical signal processing coprocessors would have much to offer. However, it is not immediately clear how such an optical coprocessor can be interfaced to the rest of the system. Setting up the two-dimensional input data serially from conventional electronic memory may largely nullify the potential advantages of such a system.[7] The limitation is actually that

---

[7] This has been referred to as the "fire hose problem" by David A. B. Miller.

of the electronic processing system as a whole, which cannot handle larger amounts of data in parallel, not of the optical coprocessor. But the bottom line is that it may not be possible to improve overall performance by simply replacing the coprocessor, because the rest of the system is not good enough to take advantage of the increased capacity and speed. (One should not exclude the possibility that in some cases the replacement might indeed prove beneficial, despite the bottleneck due to serial transfer at the interface, or there may be no interface problem because the data is already in optical form (coming from an optical memory or natural image). Nevertheless, it is likely that in most cases the capabilities of the optical subsystem will be largely underutilized due to this interface problem.)

Does this mean there is no possibility of employing optical technology short of contemplating an all optical system from scratch, which we argued was a very difficult task? Not necessarily. Some modification of the higher-level design of the system may enable a successful interface. The strength of optics in this case is that it can provide the interconnections necessary for the global flow of information, which electronics cannot. The strength of digital electronics is that it can provide complicated operations in a small space, which optics cannot. Thus, successful partitioning (or "factorization") of the overall problem to match the strengths of both technologies may lead to an architecture of the kind depicted in figure 1. The big block may represent an operation requiring regular global interconnections (implemented optically), whereas the smaller blocks represent digital electronic processing. The many smaller blocks on the left work on parts of the data independently in parallel and feed the optical subsystem in parallel, so that there is no serial bottleneck. After performing the necessary operations, the optical subsystem distributes the large array of data to the several digital processors on the right for subsequent processing.

In conclusion, we see that beneficial use of an optical subsystem may require integral redesign of the system architecture at one or more levels above that at which the optical subsystem is introduced.

5.3   Parallel memory access

Recent developments promise high-speed parallel access of huge amounts of data from silicon as well as optical memories. The former involves optical devices integrated with silicon, whereas the latter involves transmissive or reflective readout from optical storage media.
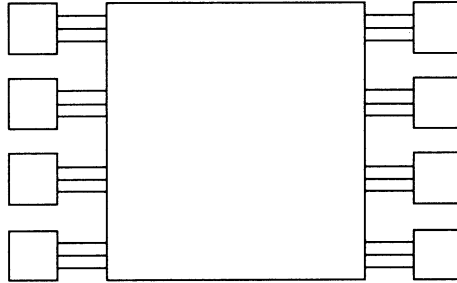
Figure 1 The large block in the middle represents global operations on the whole image. The smaller blocks on the left and right represent local operations on parts of the image.

The considerations here are similar to those in the preceding example. Although the large archival storage capacity of optical memory can be utilized without difficulty, its potential for fast parallel readout may not, unless system architectures are designed in a way that make this possible. Once again it seems that parallel architectures in the spirit of that suggested in the previous example might be useful. Alternatively, some form of optical processing may be used to distill the large amounts of data read from the optical memory, returning a small amount of information that can be handled by the electronics at the higher level. This might be useful in database applications.

5.4   Photonic digital (logical) switching

Digital photonic switches are essentially optical/optoelectronic logic circuits, often based on multi-stage switching network architectures, which enable a given pattern of connections to be established between a large number of incoming and outgoing communications lines [4].

Since the large number of incoming and outgoing lines coming from distinct locations are simply bundled together to form the "fire hose," the interface problem discussed in the image processing example does not arise in this case. This makes optical technology particularly suited to this application.

6   DISCUSSION

Optical technology can be beneficially introduced into a general purpose computer if we can come up with modifications to a system involving the use of optical components in such a way that the intrinsic advantages of the optical technology result in user-level improvements.

A particularly transparent case is to replace an electronic subsystem with an optical one of identical functionality (but perhaps completely different internal structure), in a situation where the user-level figure of merit is improved by the improvement in the external parameters of the subsystem (cost, speed, number of channels, etc.). In other words, one should identify a component or subsystem of a conventional electronic computer which constitutes a bottleneck, in the sense that replacing this component or subsystem with one that is faster (or cheaper or can handle larger amounts of data etc.), will result in the overall computer to be faster (or cheaper or able to handle larger amounts of data). (One might propose that every existing computer has a limiting component or subsystem, which is true. Nevertheless, situations in which replacement of a component or subsystem would result in substantial overall improvement may not be commonplace, since the historical evolution of design concepts and technology has a tendency to balance the various components and subsystems in a way that no component is "over-qualified" for the purpose it serves.)

Thus if such a component or subsystem is found, it would be beneficial to replace it with an optical one which exhibits improvements in the relevant characteristics. For instance, if the clock rate or power dissipation of a computing system is determined by the longest wires, and we can reduce the delay or dissipation along these wires by replacing them with optical channels, we can obtain a net improvement at the level of the overall computer. As another example, assuming that the speed of a computing system is solely determined by the memory access delay or the time it takes a coprocessor to invert a matrix, replacing these subsystems with optical ones may result in direct benefit.

The simplest special case of the above is to replace certain electrical switches or wires by optical ones without otherwise modifying the system. That is, we employ optics simply to wire up circuits designed under the conventional paradigm (low-level modification). Examples are optical backplanes, fixed free-space interconnections between circuit boards, etc. (In this case, no change is made to the existing system of levels of abstraction. A backplane or free-space interconnection system is not a subsystem in the sense of the previous paragraph.) This may not be the best way to utilize optics though, so that

disappointments in this approach should not be used to judge the potential of optics.

However, it is very difficult to do something useful with an optical technology (despite its large speed and parallelism), if it cannot be interfaced to a certain level of abstraction of an established architecture in a seamless manner. Otherwise the user-level problem description is too many levels away to make optical device arrays or the like useful for general purpose applications. Thus if the conditions for successful application of the above approaches do not exist, either because the optical technology is not directly compatible with the higher levels, or because its intrinsic advantages are buried at that level, one must modify the system architecture a few levels upwards, as discussed in conjunction with the example on page 12.

On the other hand, special purpose applications where the product can be provided in a form which can be directly used without requiring any low-level development by the user or third parties are clearly promising. Examples might be integrated optical detection planes for image preprocessing, dedicated image processing functions, "smart" optical sensors, etc.

7   CONCLUSIONS

It is clear that large arrays of very fast and low-energy optical devices integrated with established electronic technology and interconnected with free-space optics has very large computational power in the raw sense, but realizing this potential may not be so easy. The difficulty stems from the fact that a whole system of paradigms and levels of abstraction has been constructed around the capabilities and limitations of purely electronic systems, and the dominance of this system of abstractions resists the introduction of a new technology with completely different capabilities and limitations. There does not seem to be much point in trying to build an optical microprocessor, and the user-level improvements obtained by replacing the longer wires in conventional systems may be limited. On the other hand, starting with an array of smart pixels, we are too many levels of abstraction away from being able to write a program that plays chess.

Since the construction of a totally new system of paradigms and platforms is an exceedingly difficult task, it is necessary to find ways in which a manageable degree of modification of the existing system would allow net benefits at the

user level. The burden of doing this lies with those who want to promote their new technology, but it may also be possible to identify already existing computational paradigms and concepts which were until now only academic exercises but can now be implemented with optoelectronic technology.

One of the main features of the mentioned system of paradigms and levels of abstraction is that its various parts are more or less balanced in ability, in the sense that no part of the system is a bottleneck. (This is partly because more effort and resources are put into parts of the system that tend to create a bottleneck, and little effort and resources are put into those parts that are already too good compared to the rest of the system.) However, this state of affairs is dynamic; as applications change, new technologies evolve, and new ideas are introduced, it occurs that one part of the system appears as a bottleneck. Suddenly a flurry of activity begins to improve that part of the system, since any improvements in that part will automatically improve the overall machine.

A very important example that has been increasingly recognized in the past ten years is the interconnection bottleneck. Increasing use of memory, the ambition of processing large amounts of information such as with images and video, the advent of parallel computing, and purely geometrical and physical reasons are some of the factors that have contributed to the increasing importance of interconnections. The most widespread suggestion has been to replace the longer electrical interconnections with optical ones without otherwise modifying the logical architecture. Examples are optical backplanes, fixed free-space interconnections between circuit boards, etc. In this spirit, optical technology can be used to help wire up electronic circuits designed in the conventional way, by providing a large number of pinouts and high-performance long-distance connections. Although this approach certainly has a certain promise, it is not the one that we believe will bring the greatest rewards.

A more progressive approach is to replace an electronic subsystem with an optical one whose internal structure may be completely different from the electronic one it replaces. This is easier said than done, since the overall system that has been optimized with the low-performance electronic subsystem in mind may not be able to reflect the superior performance of the optical subsystem to the user level. User-level improvements would be observed only if that particular subsystem was already significantly bottlenecking the performance of the overall system, or if successful modification of the overall architecture can be made such that the optical subsystem is smoothly grafted to the overall system.

Special purpose applications in which only a few levels of abstraction are involved are excellent candidates for introducing optical technology in the short

term since the architectural and systems issues that must be tackled are less severe than those associated with general purpose systems. The difficulty here is that many such applications do not require high performance, so that already existing technologies seem to suffice. "Smart image sensing" and image processing are two related special purpose applications which seem particularly promising. Optical switching networks seems to be another. There are certain characteristics that make these applications strong candidates: (i) they involve large volumes of data (both spatially and temporally) and require global flows of information so that they strain the limits of existing systems; (ii) the format of the data and the logical organization of the processing task map naturally onto optical architectures that we know can be efficiently implemented.

Acknowledgements

REFERENCES

[1] M. R. Feldman, S. C. Esener, C. C. Guest, and S. H. Lee. Comparison between optical and electrical interconnects based on power and speed considerations. *Applied Optics*, 27:1742–1751, 1988.

[2] M. R. Feldman, C. C. Guest, T. J. Drabik, and S. C. Esener. Comparison between optical and electrical interconnects for fine grain processor arrays based on interconnect density capabilities. *Applied Optics*, 28:3820–3829, 1989.

[3] J. W. Goodman, F. J. Leonberger, S.-Y. Kung, and R. Athale. Optical interconnections for VLSI systems. *Proceedings of the IEEE*, 72:850–866, 1984.

[4] H. S. Hinton. *Introduction to Photonic Switching Fabrics*. Plenum, New York, 1993.

[5] A. V. Krishnamoorthy, P. J. Marchand, F. E. Kiamilev, and S. C. Esener. Grain-size considerations for optoelectronic multistage interconnection networks. *Applied Optics*, 31:5480–5507, 1992.

[6] D. A. B. Miller. Optics for low-energy communication inside digital processors: quantum detectors, sources and modulators as efficient impedance converters. *Optics Letters*, 14:146–148, 1989.

[7] H. Ozaktas. Using planes of optoelectronic devices interconnected by free-space optics: systems issues and application opportunities. Technical Report BLO111680-940922-57TM, AT&T Bell Laboratories, Holmdel, New Jersey, Sept. 1994. AT&T Proprietary.

[8] H. M. Ozaktas. *A Physical Approach to Communication Limits in Computation.* PhD thesis, Stanford University, Stanford, California, June 1991.

[9] H. M. Ozaktas. Towards an optimal foundation architecture for optoelectronic computing. In *Third International Conference on Massively Parallel Processing Using Optical Interconnections (MPPOI'96)*, pages 8–15. IEEE Computer Society, 1996.

[10] H. M. Ozaktas. Using planes of optoelectronic devices interconnected by free-space optics: systems issues and application opportunities. Technical Report BU-CEIS-9606, Bilkent University, Department of Computer Engineering and Information Sciences, Bilkent, Ankara, May 1996. Formerly AT&T Proprietary Report released May 7, 1996.

[11] H. M. Ozaktas and J. W. Goodman. The limitations of interconnections in providing communication between an array of points. In S. Tewksbury, editor, *Frontiers of Computing Systems Research*, volume 2, pages 61–130. Plenum Press, New York, 1991.

[12] H. M. Ozaktas and J. W. Goodman. Implications of interconnection theory for optical digital computing. *Applied Optics*, 31:5559–5567, 1992.

[13] H. M. Ozaktas and J. W. Goodman. Comparison of local and global computation and its implications for the role of optical interconnections in future nanoelectronic systems. *Optics Communications*, 100:247–258, 1993.

[14] H. M. Ozaktas and J. W. Goodman. Elements of a hybrid interconnection theory. *Applied Optics*, 33:2968–2987, 1994.