MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

Working Paper No. 258 May, 1984

ROUTING THOUGHTS

T. Poggio

Abstract: In a parallel machine with many thousands of processors the routing of information between processors is a key task, which turns out to require as much hardware and perhaps more sophistication than local computing itself. There are at least two basic engineering solutions to the routing problem: one followed by most research projects is of the "packet switching" type, that behaves as a mail service, with data carrying addresses to route the packet through the system. The other, more similar to a traditional telephone system, has connections made and broken (or enabled and disabled) as required for exchanging information. These solutions, based on silicon technology and digital electronic, may be quite different from the routing solutions used by the prototypical parallel machine – the brain.

This paper asks questions concerning routing information in parallel machines with an eye to biological wetware. It is divided in four disconnected parts, that do not contain finished results but consist of suggestions for future speculations:

1) How to make Infinity Small.

2) Routers and Brains

3) Classifying Parallel Machines

3) The Problem of Remapping

This Working Paper has been brought to you by the modern wonders of microcassette dictating equipment, through which Professor Poggio can now cough up working papers while doing something else more important.

1. How to Make Infinity Small

1.1. *Computational geometry and visual algorithms: attention as the router*

One important part of visual perception is the computation of spatial relations and spatial properties. The importance, the wide-ranging scope, and the difficulties of this class of simple computations is well-illustrated in the book *Perceptrons* by Minsky and Papert, and by the recent paper "Visual Routine" by S. Ullman. Simple relations such as the connectedness of a figure, "inside or out" (whether something is inside another figure or outside), left or right (such as the judgment one has to perform in a classical vernier task) are easily computed by the visual system, but are still a mystery from the computational point of view. It is easy enough to devise specific algorithms for each one of these tasks, but the realization of hardware capable of performing discriminations of these types over the whole visual field as humans and animals can do, essentially leads to a combinatorial explosion of numbers of cells.

This problem is reflected in what Minksy and Papert claimed was the main computational limitation of perceptron-like devices. The computational properties such as closure, connedtedness etc., require perceptrons of infinite order. In other words, each of the elementary predicates of the elementary boxes (the "predicates") used by the perceptron must be connected to each sensor in the retina. Clearly, this leads to a combinatorial explosion in the number of connections. For small retinas, however, a perceptron of infinite order is a perfectly reasonable machine. If we consider an array of, say, 100 x 100 sensors, it is not impossible to think of a perceptron of infinite order; in other words, with each predicate receiving inputs from all these 10,000 sensors.

1.2. *The Spotlight Idea*

This recent idea was proposed by A. Treisman, F. Crick and C. Koch & S. Ullman. A version of this idea that I consider here is that, while some primitive features such as color, orientation, motion and depth are computed in parallel over the whole field, most of the spatial properties and spatial relations are computed in a restricted region of the visual field in which attention, similar to a spotlight, is actually focused. The idea of a spotlight of attention has been mainly suggested in connection with the problem of establishing a conjunction of two different properties like the orientation of a bar and its color. What I am suggesting here is that a mechanism similar to the attentional spotlight may be used by the visual system to "route" a small part of the visual field to a small array of cells (a "module") dedicated to performing the specific computation required to solve the visual task of the moment.

1.3. *How to Make Infinity Small*

I argue that each one of the modules can be thought of as a perceptron of infinite order (see Poggio, 1982). In this way, we can avoid the explosion in connection numbers required by a perceptron machinery operating over the whole of the retina. Assume, for instance, that the output of the retina is about one million fibers, and that in another area of the brain, there is a set of small perceptrons, each one (in general) of infinite order, but looking at a small set of inputs (say, an array of 100 x 100). Each one of these perceptrons could be trained as suggested by several existing perceptron learning theorems. The attentional searchlight would select a specific small spot of the visual field for one of these perceptrons. Assume for instance that the task to be performed is to decide whether a certain segment is to the left or to the right of a lower segment—the typical vernier task. As a result of training, a small perceptron capable of performing this task for certain stimuli would be created with, perhaps, an input of 100 x 100 fibers. The attentional searchlight will select some region of the visual field and *connect* it to this specific module, which is, in turn, selected by a top level processor among all the possible modules. In this way, the attentional searchlight can give us the full power of an infinite order perceptron without the explosion of connections that are intrinsic to it. The mechanism by which the searchlight selects a specific spot in the visual field does not have to concern us here, but a mechanism of the winner-take-all type may well account for most of the needs. (See Koch & Ullman.) A scheme of this type cannot, of course, solve by itself global problems such as connectedness of a large figure. In order to solve these problems, the path of the searchlight must be controlled and the result of the operations performed during this trajectory must be accumulated and further processed by some more central module. (See, for instance, Ullman, 1983.)

1.4. *Summary*

We suggested that spatial relationships and various problems in computational geometry are solved by perceptron-like devices. These perceptrons have, in general, infinite order, but the retina they can access is small. The mechanism responsible for routing a small part of the visual field to one of these devices is the attentional spotlight. The emerging picture, then, is one of many small perceptrons with a *routing device* which directs their connections to a selected and small part of the visual field depending on the task at hand and on the location in the visual field of what is relevant. A bold physiological prediction is that area VII may be a part in the cortex where these several perceptron-like devices are located. If this is the case, the large receptive fields measured in area VII may just reflect a random-like walk of the attentional spotlight, and not the precise pattern of connections at any instant in time from these cells to the visual field. The prediction is that at any instant of time, the receptive fields of these neurons is quite small, but that since the virtual connections of these modules will change with the trajectory of the spotlight, the average receptive fields over some time can be quite large.

It would be interesting to simulate on a toy program some of the features of this idea. (This may be especially relevant in the context of the connection machine.) Small perceptron-like routines residing in some part of the connection machine will be accessed through the router system, and will be applied to different parts of the visual image. The architecture of such a machine would be quite interesting: a number of elementary features could be computed continuously over the whole image for each of the individual processors such as motion, depth, orientation and color, but only a few, more specialized operations will be performed on small parts of the image at any one time. Two selection mechanisms are required: one for choosing which module to apply, i.e., which perceptron to apply; the other for deciding where to focus the operation of this perceptron. One particular organization for visual computation that comes to mind is the following. Motion is detected. Attention is focused on the most relevant point of the visual field. A perceptron is activated to check whether the motion is translation or not. If not, another perceptron is called in to check whether the motion is rigid or not. If one of these modules called in sequentially answers positively, then another sequence of specific operations is started, for instance, to compute structure from motion. The usual structure of hypothesis production and hypothesis verification may be implemented very effectively in such an architecture based on attentional mechanism and small perceptrons. One of the attractions of this scheme is the simplicity with which new routines, i.e., new perceptrons, could be synthesized and old one modified by training.

2. Routers and Brains

In the last few years, new architectures for parallel computers have been developed. The most advanced machines try to overcome the so-called von Neumann bottleneck by having hundreds of thousands of individual processors computing in parallel and communicating with each other. Although these designs are very new, it is becoming increasingly clear that the critical problem is not the computational unit, but the communication network.

In most of the existing designs, a large part of the hardware on-chip is indeed dedicated to routing messages between the processors. Different routing connections have been developed. The different possibilities range from the *cross bar* to the new *fat trees* developed by Charles Leiserson.

Clearly, much theoretical work has still to be done in order to characterize optimal routing schemes. For instance, most of the existing routers assume that the algorithms require essentially uniformly distributed connections among all processors. In many practical cases, however, one would expect that algorithms can be devised so that neighboring connections will be used much more frequently than long distance connections (somewhat similar to what is happening in the telephone exchange network).[1] In any case, from the point of view of theoretical neuroscience, it is very suggestive that the routing problem has such a critical role and requires such a large part of the hardware of the new parallel machines. The obvious question is whether this routing problem has the same importance in the brain, which, after all, has been very often used as the typical example of a parallel architecture.

It would be quite intriguing if a large percent of the computational hardware and of the biophysical mechanisms of neurons and synapses would be dedicated, not to raw processing of signals, but to their routing from cells to cells and from areas of the brain to other areas of the brain. This is an intriguing conjecture that may trigger exciting speculations and perhaps even new experimental work in the neurosciences.

3. A Classification of Parallel Machines

The ideas discussed in this section come from comments of Shimon Ullman on the first part of this working paper.

If routing is so important in a parallel machine, it may be interesting to explore the possibility that one could classify different types of parallel machines on the basis of the routing scheme that is used. A specific possibility is the following: Consider the architecture described in a previous section of this working paper, "How to Make Infinity Small"; information is routed from the input part of the computer to a computing part, where many different boxes containing specialized algorithm and dedicated hardware is. The other extreme possibility is the visual routine architecture corresponding to the ideas described by Ullman in his "Visual Routine" paper. In this last case, very simple routines are executed; although each one is unable to do much, an intelligent combination of several of them may solve very complex problems. In the first case, on the other hand, each of the "boxes" is a powerful, dedicated computing device.

The two schemes are also different in that in the visual routines scheme, programs are so to speak, "sent back" to the input where they operate. In the infinite perceptrons case, information from the input region is sent to this highly specialized and powerful machines. The question, which remains somewhat vague at this point, is whether one could develop a useful classification of parallel architectures, say, for vision, based on distinctions of this type.

[1] In the Connection machine, for instance, neighboring connections are also present in addition to the long distance routing network.

4. The Problem of Remapping

Motter and Poggio have recently obtained intriguing evidence that some kind of remapping of the visual field takes place in cortical neurons in VI during small eye movements. When a monkey fixates a target, the position of the maximum response of a visual cell is, according to their measurements, independent of small deviations of the eye position from the fixation target. The amount of deviation is around a maximum of 20 minutes of arc with a receptive field that can be as small as 3 minutes or less. When larger eye movements take place, an apparent reset of the whole system occurs and the receptive field changes accordingly. It seems, therefore, unavoidable to conclude that during *small* movements of the eyes around a fixation point, the receptive field of cortical neurons are continuously remapped in such a way that their position is stationary in the cortex. The evidence demonstrated by Motter and Poggio is strong; however, given the revolutionary nature of these findings, a critical experiment would be highly desirable. The demonstration that the visual field moves in the retina or perhaps even in the LGN during small eye movements while remaining stationary in the cortex would be a particular dramatic demonstration of the findings.[2]

4.1. *Why Remapping is Either Wrong or Revolutionary*

Some kind of remapping is the obvious explanation for the apparent constancy of the visual world when we move our eyes. This is especially true from the "homunculus" point of view (which is looking at the cortex as a television screen). Note that remapping in a topographical area is not strictly necessary for explaining the stability of the visual world. We simply need to know that nothing strange occurs in the outside world—such as, for instance, relative motion between objects—in order to perceive stability of the visual world. Needless to say, if remapping occurs at the level of VI or even of the LGN, the traditional picture that we have of the visual cortex (for instance, from the work of Hubel and Wiesel) will change dramatically. Furthermore, it is very difficult to conceive which circuitry and which biophysical mechanisms should be involved to explain how remapping can be done. It is clearly very easy to image how to remap the whole topographical visual field in terms of computer's shift registers, for instance. It is, however, very difficult to describe how neurons with the standard properties that are known could be made to perform the precise remapping required by Motter and Poggio's experiments. This note asks the question of how to build nervous hardware for the remapping question.

4.2. *How to Perform Remapping and Routing with Neurons*

The problem of remapping is a special instance of a more general problem with which neurons are confronted. How can information be routed in controlable ways to different

[2]Imprecision in the measurements of the eye movements and possible interference from head movements, although controlled by Motter and Poggio, can still be advocated to avoid the conclusion that remapping takes place in the visual cortex.

destinations? The corresponding operation in today's computer is based on the capability of addressing messages to different locations. Neurons do not seem to have addresses or the capability of sending addressable messages. How then can the visual map be shifted exactly to correct for eye movements? This is routing of the packet switching type. Another possibility is to establish and break connections as needed, similar to the standard telephone exchange network. In this case, connections may physically exist, but only some are activated (or *enabled* at any time. We outline here three different possibilities.

In the first case, we assume that information from eye muscles is very reliable about the precise instantaneous position of the eyes. This information may enable different subsets of the widely spread connections between the cortical cells and the photoreceptors (by, of course, several intermediary neurons). This mechanism requires connections extending quite a long way across the cortex; it also requires a separate synaptic mechanism of the type devised by Koch and Sherman to switch off parts of the receptive fields, perhaps of the X system.

The second mechanism does not rely on measuring eye position, but relies on measuring the distance of a given point from the fixation target and using this measure to perform the same operation as before. This scheme is more complicated, and it requires feedback from higher centers where this distance is possibly computed.

The third possibility is most intriguing. It is possible that in the first moments of the experiment, the response of the cell elicited when the eye fixates on the target becomes associated with features of that particular target within a few minutes' distance from the center of the receptive field of the cell. When the stimulus occurs again, the response learned during the previous presentation is "retrieved". This requires a very fast learning process that could be reset by a number of higher level processes; for instance by the detection of relative motion in the image. It is unclear whether this imprinting of the image into the cortex may be done under more general and less repetitive situations that the one involved in the experiment by Motter and Poggio. This last hypothesis would require the existence of quickly changing synapses (perhaps Marsburg synapses such as the synapses Crick (1984) has proposed).

4.3. *Conclusions*

All these explanations are definitely in terms of standard properties of neurons. It seems worthwhile to us to make them explicit, first for stressing that it would indeed by very important to establish the existence of remapping, and second, for stimulating some more thoughts about how remapping could be done. The routing of information through different channels without standard address capabilities such as computers have, is an important problem that the nervous system must solve over and over again. The phenomenon of remapping in VI in addition to its intrinsic importance, may also represent a clear example of how routing can be done without addressing by the nervous system.

5. Questions to be answered

1.1) Do theory and computer experiments on machines based on "small", infinite order perceptrons.

1.2) Psychophysics similar to Barlow's and Watson's to test the idea?

2.1) Routing in the brain. Importance, anatomy, possible examples.

2.2) What are the biophysical mevhanisms used by the brain's routers?

3.1) Develop theory based on routing scheme. Add idea of locality.

4.1) Model biophysical solution to Motter and Poggio's experiments.

Acknowledgments: Jim Mahoney was patient enough to read this blurb. C. K. was invaluable. ALY wants everything to be in the striatum. Carol was unusually cooperative and, as usual, very good.

References

Crick,F.: The function of the thalamic reticular complex: the searchlight hypothesis. *Proc. Natl. Acad. Sci. USA*, in press (1984)

Koch, C. "A theoretical analysis of the electrical properties of a X-cell in the cat's lateral geniculate body: Does the spine-triad circuit subserve visual attention?", *Artificial Intelligence Lab. Memo*, **No. XXX**, MIT, Cambridge, MA 1984.

Koch, C. and Ullman, S., "Selecting one among the many: A simple network implementing shifts in selective visual attention", *Artificial Intelligence Lab. Memo*, **No. 770**, MIT, Cambridge, MA 1984.

Minsky,M., & Papert,S.: *Perceptrons*, Cambridge, Massachusetts: MIT press (1969)

Motter, B.C. and Poggio, G.F., "Binocular fixation in the rhesus monkey: Spatial and temporal characteristics", *Exp. Brain Res.*, **54**, 304–314, 1984.

Ullman, S., "Visual routines", *Artificial Intelligence Lab. Memo*, **No. 723**, MIT, Cambridge, MA 1983.