



# Toward Identifying the Systems-Level Primitives of Cortex by In-Circuit Testing

Leslie G. Valiant\*

Harvard University, Cambridge, MA, United States

**Keywords:** computational primitives, memory, learning, chunking, association, quantitative analysis, systems-level

## SUMMARY

The hypothesis considered here is that cognition is based on a small set of systems-level computational primitives that are defined at a level higher than single neurons. For one such set of primitives, quantitative feasibility has been demonstrated by analysis and computer simulation. In this note we point out that emerging technologies for stimulation and recording are making it possible to test directly whether cortex is capable of performing this particular set. Such an experimental program would therefore provide a potential avenue toward uncovering the actual intermediate-level primitives used by cortex.

## INTRODUCTION

In both biology and computer science the virtue of modularity in complex systems is widely understood. In contrast, the idea that complex systems need to have many levels of description, while well appreciated in computer science, has been much less discussed in biology. We believe that for a computational theory of cortex to be useful, it would need to give an account of algorithmic tasks at levels intermediate between those of neurons and behavior. Just as we cannot understand the workings of a cell phone by studying its operations only at the single bit level, we cannot expect to understand the brain at the single neuron level.

The suggestion we explore here is that cognitive computation in cortex is built on a small collection of base primitives at the intermediate systems level. This contrasts with Hebbian plasticity (Hebb, 1949), which is a single primitive at the single neuron level. Here we discuss one candidate for such a collection of base primitives, one distinguished, perhaps uniquely, by the existence of demonstrations, by analysis and computer simulations, that it offers a scheme in which the capacity of cortex for basic cognitive tasks can be given a quantitative explanation (Valiant, 1994, 2005, 2006; Feldman and Valiant, 2009) in terms of the resource constraints on neuron numbers, connection numbers, and synaptic strengths. This collection consists of: association, memorization, inductive learning, and hierarchical memory assignment, all suitably defined. The purpose of this note is to point out that because of technological advances over the last decades, especially in optogenetics (Miyawaki et al., 1997; Boyden et al., 2005; Deisseroth et al., 2006; Shemesh et al., 2017), it is now becoming feasible to test systematically whether cortex can indeed perform these particular primitives.

These four primitives are defined below in terms of their actions on *randsets*, random sets of neurons of a certain size  $r$ . The nature of the proposed experiments is analogous to “in-circuit testing” of a traditional electronic printed circuit board, where instead of testing the input-output behavior of the whole board, the testing is done on smaller parts, using only probes. In the proposed experiments there will be no sensory stimuli or behavioral responses—the experiments will consist

## OPEN ACCESS

### Edited by:

Michael M. Halassa,  
Massachusetts Institute of  
Technology, United States

### Reviewed by:

Adam Ponzi,  
Okinawa Institute of Science and  
Technology Graduate University,  
Japan

### \*Correspondence:

Leslie G. Valiant  
valiant@seas.harvard.edu

Received: 24 July 2018

Accepted: 31 October 2018

Published: 20 November 2018

### Citation:

Valiant LG (2018) Toward Identifying  
the Systems-Level Primitives of Cortex  
by In-Circuit Testing.  
Front. Neural Circuits 12:104.  
doi: 10.3389/fncir.2018.00104

entirely of stimulating and recording from appropriate sets of neurons. The goal is to find out what changes in computational function in a subcircuit of cortex can be achieved by specific neural stimulation. Existing experimental evidence that the firing patterns of cortical neurons can be altered by suitable behavioral training (Jackson et al., 2006; Rebesco et al., 2010) is relevant and encouraging, but does not directly address the question of which computational primitives are being executed. Behaviors prompted by sensory inputs cause all kinds of activity in the brain extraneous to the execution of any one primitive. The in-circuit methodology is suggested to isolate the operation of a subcircuit as much as possible from such interference.

The goal therefore is to test, by direct neural stimulation and recording, whether cortex is capable of performing certain basic computational tasks. The paradigmatic first such task is “association of a set  $A$  to set  $B$ ,” abbreviated as  $A \rightarrow B$ , which here is defined as follows: for neuron sets  $A$  and  $B$ , each of  $r$  neurons, we want to induce the new functionality that, after the training, whenever set  $A$  is stimulated the newly modified circuit will cause set  $B$  to become active also. Training here would involve stimulating  $A$  and  $B$  according to some timed protocol. Testing would involve stimulating  $A$  again and now recording from the set  $B$  that had been stimulated in training. This particular task may be viewed as a direct systems level analog of Hebbian plasticity. Note that the hypothesis that such operations work for random sets of neurons, most of which have no direct pairwise connections, is a strong one. Yet it is easily testable since it is sufficient to stimulate and record from randomly chosen sets of neurons.

Our notion of randsets is not the same as the notion of assemblies in the sense of Hebb (1949). That traditional notion of assemblies is associated with the idea that the neurons in these assemblies are better connected to each other than to other neurons (e.g., Carillo-Reid et al., 2017). Randsets account for the claimed computational capabilities simply by virtue of being randomly chosen: analysis shows that the random choice assumption is sufficient to guarantee that the claimed algorithms succeed with overwhelming probability. Note that there is no comparable theory that shows that the high interconnectivity assumption of traditional assemblies, without the randomness of the sets, provides any comparable computational capabilities. Fortunately, also, it is the randset theory that appears to be one that is more readily testable.

An experiment relevant to what we are contemplating here is reported in Seeman et al. (2017), where the setting up of such associations was attempted by electronic stimulation, but did not succeed reliably. The experiment was done for one value of  $r$ . Possible interpretations are that the value of  $r$  used there was too small, or that it was not a part of cortex that performs this function. We are suggesting a systematic exploration of the relevant parameter space, and in particular the set size  $r$ . Another relevant result, one obtained by optogenetic means, is by Carrillo-Reid et al. (2016), which can be viewed as also an example of “in-circuit” testing, but the task considered there was the single one of pattern completion (e.g., Hopfield, 1982).

## THE FOUR TASKS AND THEIR POTENTIAL VALIDATION

Each instance of the tasks described here can be induced by an associated training protocol, and is realized as local changes in the neurons within the given randomly connected network, giving rise to a subcircuit having the functionality of that task instance. Computer simulations (Feldman and Valiant, 2009) show that thousands of instances of such tasks can be acquired in succession with realistic network parameters without having the continued effectiveness of the ones acquired earlier degraded by later ones. We note that having multiple task types, which between them may be adequate to form a basis for cognition, is a much more exacting requirement than single-task models such as traditional pattern completion. The longer term motivation is that of identifying a “cognitively adequate” set of primitives, perhaps our four forming the core, and showing that higher level cognitive operations can be efficiently implemented in terms of these.

We now detail the four tasks that we are suggesting as the core, and for which we are here suggesting validation by in-circuit testing. An example of some more detailed specifications still, for one incarnation, is described by Feldman and Valiant (2009).

### Association

For sets  $A, B$  train so that result is: If in future  $A$  is stimulated then  $B$  (but not random other neurons) will become active.

### Supervised Memorization of Conjunctions

For sets  $A, B, C$  train so that the result is: If in future  $A$  and  $B$  are stimulated then  $C$  will become active also (but not if just one of  $A, B$  is stimulated).

### Inductive Learning of Simple Threshold Functions

For sets  $A, B, C, D$  (say) train so that the resulting circuit generalizes a classification at target  $A$ , beyond the examples seen, according to a linear separator on the variables  $B, C, D$ . In particular, if in future some subset of  $B, C, D$  is stimulated then the target  $A$  will become active according to whether some linear separator consistent with the examples holds. For example such a linear separator may be  $B + C + 2D \geq 2$ , where now these variables have a 1/0 interpretation corresponding to whether the randsets are active or not.

### Hierarchical Memory Formation

For sets  $A, B$ , train so that the result is: a set  $C$  (not specified by the experimenter) has been allocated by the computation internally so that the effect “supervised memorization of conjunctions” is achieved for  $A, B$  and that, now unknown,  $C$ . This realizes the fundamental psychological task of “chunking,” where a new compound concept, here  $A \& B$ , becomes equal citizen in the circuit with earlier ones.

While the first three of these tasks can be tested directly by in-circuit testing, the last task, where one does not have direct access to  $C$ , would require indirect inference, perhaps from the timing delays introduced by  $C$  downstream.

Verifying *how* these tasks are realized in cortex, if indeed they are, is a further area of challenge. The suggestion has been made (Valiant, 2012) that the hippocampal system is responsible for determining the identity of the randset  $C$  in hierarchical memory formation, for given  $A$  and  $B$ . Such a statement, which assigns responsibility for performing a particular task to a particular brain area, may be easier to test.

## GENERAL CONSIDERATIONS

The experiments we are proposing would establish that the brain can implement the primitives that we are hypothesizing. However, the biological implementation could be realized in any number of ways. Experimentation will be needed to determine which one, if any. In this sense what we are proposing is a broad research program, rather than a single experiment.

### How Large Sets?

In the analysis the parameter  $r$  is dependent on the number of neurons in the system, the number of connections, the synaptic strengths, and the algorithms used for realizing the tasks. For the association task, for example, having  $r$  too small will fail to train the target  $B$  as desired, while having  $r$  too large will train too many spurious neurons in addition. Hence, we expect that considerable experimentation may be needed to find the right value of  $r$  in any one cortical area. Note that in this case of associations, the size of the source set  $A$  appears to be critical, but this is the easier set in that it only needs stimulation. The size of the target set  $B$ , which has to be both stimulated and recorded from, may be much less critical.

### Which Brain Areas?

To test our theory it is most logical to perform the experiment in a cortical area where the corresponding task, such as association, is implemented in biology. There is some uncertainty as to which parts of cortex perform the various such tasks, and experimentation will be needed to find them. The hypothesis is that the claimed primitives are the ones used by behaving animals in the presence of background brain activity. For the experiments we have in mind behaving mammals.

## REFERENCES

- Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G., and Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* 8, 1263–1268. doi: 10.1038/nn1525
- Carrillo-Reid, L., Yang, W., Miller, J. E., Peterka, D. S., and Yuste, R. (2017). Imaging and optically manipulating neuronal ensembles. *Ann. Rev. Biophys.* 46, 271–293. doi: 10.1146/annurev-biophys-070816-033647
- Carrillo-Reid, L., Yang, W., Bando, Y., Peterka, D. S., and Yuste, R. (2016). Imprinting and recalling cortical ensembles. *Science* 353, 691–694. doi: 10.1126/science.aaf7560
- Deisseroth, K., Feng, G., Majewska, A. K., Miesenbock, G., Ting, A., and Schnitzer, M. J. (2006). Next-generation optical technologies for illuminating genetically targeted brain circuits. *J. Neurosci.* 26, 10380–10386. doi: 10.1523/JNEUROSCI.3863-06.2006

## How Sets Chosen?

Since the hypothesis requires that the experiment work for randomly chosen sets of a certain size in the appropriate brain areas, it is reasonable to go with whatever bias the experimental technique used imposes. For example, electrical stimulation by electrodes may stimulate cells within a certain proximity, while optogenetic methods may do so for a larger region. Neurons in particular layers may need to be found. The various sets  $A$ ,  $B$ ,  $C$ , etc., may be in the same cortical area, or in different areas.

## What Timing Protocols for Training?

Timing may be critical. When stimulating a pair of sets  $A$ ,  $B$  for association we need an asymmetric protocol, such as  $B$  being stimulated with 1 ms delay after  $A$ .

## Comparisons?

Comparative studies for different cortical areas and different species, including primates, would be of great interest.

## CONCLUSION

The hypothesis to be tested is that cortex is able to perform a specific set of tasks on random sets of neurons. The new opportunity is offered by the remarkable recent advances in stimulation and recording technologies. The experiments would determine only *whether* appropriate cortices have the ability for performing these tasks. However, positive results would demonstrate that the brain has impressive computational capabilities that are relevant to cognitive computation. These systems-level capabilities, if present, are unlikely to have arisen by accident. It would be a remarkable coincidence if the brain had these capabilities but did not use them.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

## ACKNOWLEDGMENTS

This work was supported in part by a grant from the National Science Foundation CCF 1509178.

- Feldman, V., and Valiant, L. G. (2009). Experience-induced neural circuits that achieve high capacity. *Neural Comput.* 21, 2715–2754. doi: 10.1162/neco.2009.08-08-851
- Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York, NY: Wiley and Sons.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 9, 2554–2558. doi: 10.1073/pnas.79.8.2554
- Jackson, A., Mavoori, J., and Fetz, E. E. (2006). Long-term motor cortex plasticity induced by an electronic neural implant. *Nature* 444, 56–60. doi: 10.1038/nature05226
- Miyawaki, A., Llopis, J., Heim, R., McCaffery, J. M., Adams, J. A., Ikura, M., et al. (1997). Fluorescent indicators for  $Ca^{2+}$  based on green fluorescent proteins and calmodulin. *Nature* 388, 882–887. doi: 10.1038/42264

- Rebesco, J. M., Stevenson, I. H., Körding, K. P., Solla, S. A., and Miller, L. E. (2010). Rewiring neural interactions by micro-stimulation. *Front. Syst. Neurosci.* 4:39. doi: 10.3389/fnsys.2010.00039
- Seeman, S. C., Mogen, B. J., Fetz, E. E., and Perlmutter, S. I. (2017). Paired stimulation for spike-timing-dependent plasticity in primate sensorimotor cortex. *J. Neurosci.* 37, 1935–1949. doi: 10.1523/JNEUROSCI.2046-16.2017
- Shemesh, O. A., Tanese, D., Zampini, V., Linghu, C., Piatkevich, K., Ronzitti, E., et al. (2017). Temporally precise single-cell resolution optogenetics. *Nat. Neurosci.* 20, 1796–1806. doi: 10.1038/s41593-017-0018-8
- Valiant, L. G. (1994). *Circuits of the Mind*. New York, NY: Oxford University Press.
- Valiant, L. G. (2005). Memorization and association on a realistic neural model. *Neural Comput.* 17, 527–555. doi: 10.1162/0899766053019890
- Valiant, L. G. (2006). A quantitative theory of neural computation. *Biol. Cybernetics* 95, 205–211. doi: 10.1007/s00422-006-0079-3
- Valiant, L. G. (2012). The hippocampus as a stable memory allocator for cortex. *Neural Comput.* 24, 2873–2899. doi: 10.1162/NECO\_a\_00357

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Valiant. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.