Analyzing Composability in a Sparse Encoding Model of Memorization and Association

Jacob Beal
MIT CSAIL
Cambridge, MA, 02139
jakebeal@mit.edu

Thomas F. Knight, Jr
MIT CSAIL
Cambridge, MA, 02139
tk@csail.mit.edu

***Abstract*—A key question in neuroscience is how memorization and association are supported by the mammalian cortex. One possible model, proposed by Valiant[10], uses sparse encodings in a sparse random graph, but the composability of operations in this model (e.g. an association triggering another association) has not previously been evaluated. We evaluate composability by measuring the size of "items" produced by memorization and the propagation of signals through the "circuits" created by memorization and association. While the association operation is sound, the memorization operation produces "items" with unstable size and produces circuits that are extremely sensitive to noise. We therefore amend the model, introducing an association stage into memorization. The amended model preserves and strengthens the sparse encoding hypothesis and invites further characterization of properties such as capacity and interference.**

I. Introduction

A key question in neuroscience is how memorization and association are supported by the mammalian cortex. The question is complicated by the measured number, degree, and synaptic strength of cortical neurons. Neurons appear to be sparsely connected: while the number of neurons in mouse cortex is estimated to be 1.6×10^7 and the number of neurons in human cortex is estimated to be approximately 10^{10}[4], their degree—the number of neurons with which each neuron synapses—is estimated to be much smaller, approximately 7800 in mouse cortex and 24,000-80,000 in human cortex[1]. At the same time, the average strength of synapses appears to be quite weak, with each estimated to effectively contribute a fraction in the range 0.003 to 0.2 of the firing threshold[1]. While some significantly stronger synapses have been recorded ([9], [7], [2]), weak synapses appear to predominate and some neural systems may be dependent entirely on weak synapses.

A model of memorization and association, using only weak synapses and consistent with these parameters, has been proposed by Valiant[10]. In this model, interconnecting cortical neurons are modelled as a sparse random graph and "items" to be manipulated are represented as sparse subsets of graph nodes (decisions consistent with a long history of work on neural networks, including [3], [6], [5], [1], [8], and many others). In this model, an "item" represented by a set of graph nodes is considered to be "recognized" when at least half of the nodes in the set are firing, so nodes may be used in more than one item, so long as the overlap between items is small.

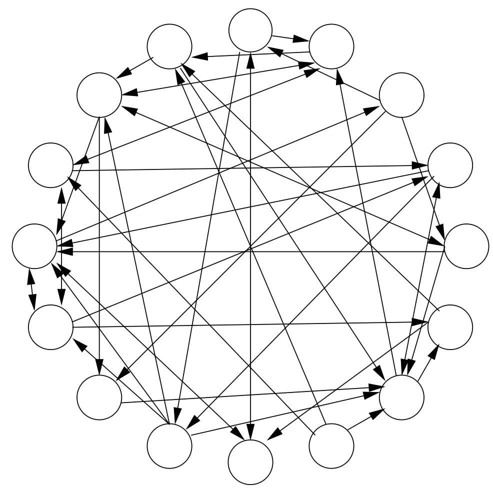


Fig. 1. The sparse encoding model proposed in [10] models a portion of cortex as a sparse random graph with directed edges, where each edge has a weight representing its synaptic strength and each node fires when the incoming edges from firing nodes sum to a high enough weight.

Memorization is the joining of two items, A and B, to create a new item C, such that C is recognized if and only if both and A and B are recognized (an "AND" relationship). Association is the linking of two items, D and E, such that whenever D is recognized, E is recognized also (an "IF" relationship). In neither case should there be side-effects or interference from other memorization or association relationships. The original proposal includes two algorithms, JOIN and LINK, that create "circuits" implementing memorization and association, respectively.

While these algorithms and the "circuits" they produce are evaluated in isolation in the original proposal, their composition (e.g. an association triggering another association) has not previously been evaluated. We empirically evaluate the model's performance on two aspects of composition: the variation of item size during repeated memorization and the impact of noise on signals propagating through memorization and association "circuits." The evaluation shows that association is sound, but that memorization is extremely sensitive to size variation during construction and produces circuits in which signals degrade badly in the presence of any noise. We therefore amend the model to abolish these sensitivities by incorporating an association stage into memorization.

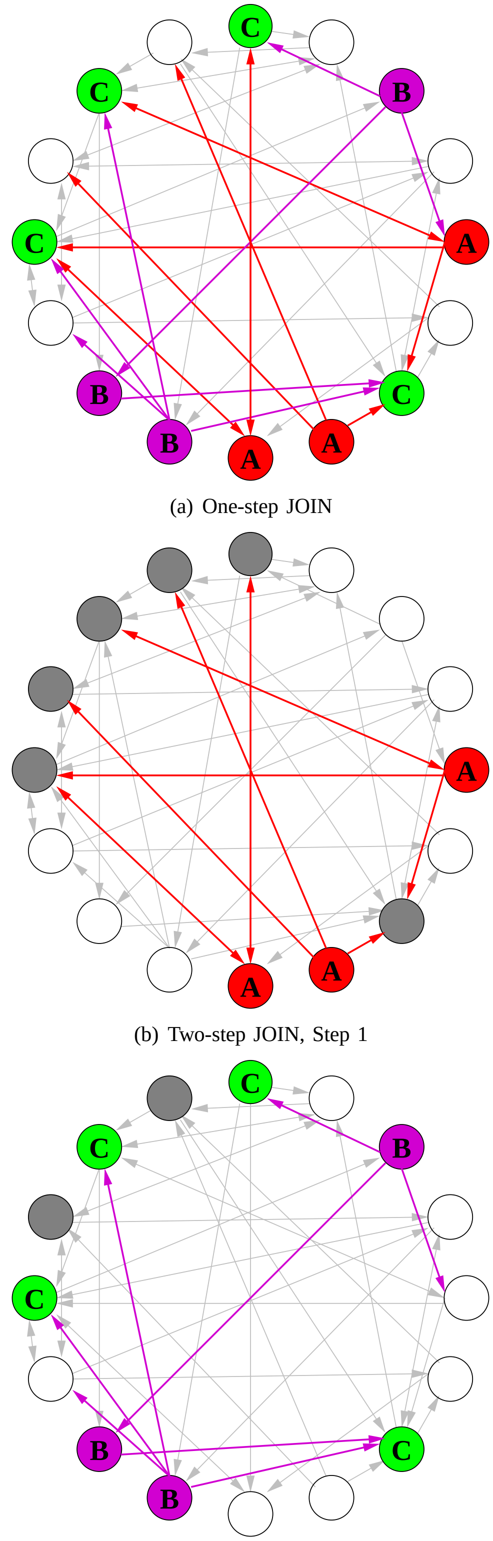
II. Memorization and Association in Sparse Random Graphs

The sparse encoding model proposed by Valiant[10] consists of four modular components: a model of cortex as a general or bipartite sparse random graph, representation of "items" as disjoint or shared sparse sets of graph nodes, one-step and two-step JOIN algorithms that create composite items, and a LINK algorithm that associates one item with another. This section reviews the model from [10], noting which variants are used and our implementation decisions where the original model is underspecified.

A portion of cortex—either a single brain region or several interconnected regions—is modelled as a sparse random graph. The graph may be a general graph with n nodes (Figure 1) or a bipartite graph where each component has n nodes. Each node has an expected degree d and has probability $p = d/n$ of connecting to each other node in the network with a directed edge. At high n and low p, the general graph and bipartite graph behave similarly, so we will treat only the case of a general graph, for simplicity.

Each edge has a weight associated with it, representing synaptic strength. Each graph node is either firing or inactive, and contains a simple finite state machine that determines its behavior. The level of stimulus at a node is equal to the sum of weights on incoming edges. When this stimulus exceeds a threshold (normalized to 1), the node fires unless its current state suppresses firing. Memorization and association use different base synaptic strengths, $1/k_m$ and $1/k_a$. For pragmatic reasons, our empirical evaluations use one population of edges with multiple weights rather than multiple populations of edges, but these are equivalent because only one set is ever in use at any time. Timing is left vague in [10], so we calculate firing in steps, where the stimulus at a node in a given step comes from the nodes firing in the previous step.

Memorization and association operate on "items," which might represent concepts or other fragments of mental state. An item is represented by a set of r graph nodes, where $r << n$, and we assume the existence of a mechanism that creates items on demand by selecting a random set of nodes. The representation of items may be either disjoint, meaning that each node is included in zero or one items, or shared, meaning that each node may be included in many items. In this paper, we consider only the shared representation, because many more items can be allocated than in the disjoint representation. The disjoint representation can only allocate a maximum of n/d items, and memorization fails when only a small fraction of that limit has been allocated (due to the sensitivity discussed in the next section). The shared representation also has the advantage of providing a possible explanation for the surprising ease with which stimuli can be found that produce selective responses in individual neurons: if a neuron participates in many items, it is much easier to find some item that causes it to respond.

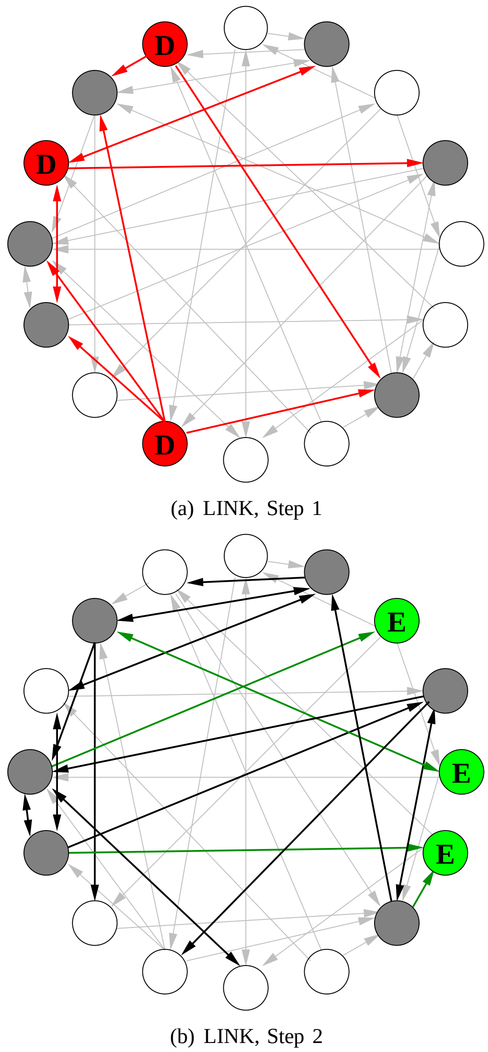


(a) One-step JOIN

(b) Two-step JOIN, Step 1

(c) Two-step JOIN, Step 2

Fig. 2. The JOIN algorithm uses existing items A and B to create an item C that is recognized if and only if A and B are recognized. In one-step JOIN, A and B are triggered to fire simultaneously and C is the nodes they stimulate to fire. Two-step, using twice the edge weight, first triggers A, moving nodes that would fire to an intermediate state, then triggers B to fire and C is the intermediate-state nodes that fire.



(a) LINK, Step 1

(b) LINK, Step 2

Fig. 3. The LINK algorithm connects existing items D and E so that if D is recognized, then E is recognized as well. D is triggered to fire and the firing is propagated for two steps. In the first step, all edges have weight $1/k_a$, and in the second step all edges initially have weight 0. LINK works by raising the second-step weight to $1/k_a$ on edges that arrive at E from firing nodes.

If a large enough fraction of an item's nodes are firing, then the item is considered to be "recognized;" if few are firing then the item is considered to be "not recognized." In the original model, the threshold is fixed at 50%; in this paper we generalize this by allowing an "invalid" range that separates "recognized" and "not recognized" states, just as digital circuits have an invalid range of voltages between those that represent binary "1" and those that represent binary "0."

Memorization is implemented using the JOIN algorithm, which uses existing items A and B to create a new item C that is recognized if and only if A and B are recognized. JOIN comes in one-step and two-step variants (we ignore the disjoint representation variants for JOIN and LINK), both of which work by triggering A and B to fire, then identifying C as the set of nodes that are well stimulated by both A and B (Figure 2). In both cases, all edges have weight $1/k_m$, which is set to give C an expected size of r nodes. The edge weights are never adjusted. One-step JOIN uses a k_m twice the size of two-step JOIN. In one-step JOIN, A and B are triggered simultaneously, and C is the set of nodes that fire. In two-step JOIN, first A is triggered, and all nodes that should fire instead advance to an intermediate state. Next, B is triggered, only nodes in the intermediate state are allowed to fire, and C is the set of nodes that do fire. Execution of these "circuits" is the same as construction, but using the current firing level for A and B rather than triggering them, and looking at C for the outcome. Two important notes: first a symbol cannot be joined with itself, and second, due the statistics of the random graph, executing more than one JOIN at the same time is likely to cause every node in the network to fire.

Association is implemented by the LINK algorithm, which connects existing items D and E, so that if D is recognized then E is recognized as well. LINK executes in two steps, and does adjust weights (Figure 3). In the first step, D is triggered to fire and every edge has weight $1/k_a$. The parameter k_a is set so that at the second step nearly all nodes will have at least k_a edges incoming from firing nodes. The weights for the second step are all initially 0, but LINK raises them to $1/k_a$ on edges that arrive at E from firing nodes. Thus, when D is recognized, precisely those items that have been connected to it by LINK will be recognized two steps later.

III. Sensitivity of Memorization to Item Size

Chained memorization circuits are an area of concern for composition because the size of the new item created by JOIN depends on the size of the two items being joined, yet also varies due to the randomness of the graph. The original proposal notes that this variation can be expected to be on the order of $\sqrt{r}$, which is only a small variation in size. When items produced by JOIN are themselves used in a memorization, however, even a small variation may have a large impact. The size of the item created is determined by the size of the upper tail of a random distribution, and the amount of the tail above the firing threshold is extremely sensitive to variations in the mean of the distribution.

We demonstrate the importance of this effect by evaluating the degree of sensitivity empirically. Both one-step JOIN and two-step JOIN are evaluated, using parameters from [10]: $n = 100,000$, $d = 512$, and the corresponding pairs $k_m = 32$, $r = 2134$ for one-step JOIN and $k_m = 16$, $r = 2338$ for two-step JOIN. Figure 4 shows the size of the new item created from a pair of items ranging from $0.9r$ to $1.1r$ in steps of $0.01r$, with 100 samples for each size (10 times on 10 networks). Small variations in the size of the initial items are greatly amplified in the size of the item created by the JOIN. With these parameters, both one-step and two-step JOIN amplify small size variations by approximately one order of magnitude.

The high sensitivity of JOIN to the size of the initial items means that chaining together even a small number of JOIN operations is unstable, and that even a few iterations leads to representations that contain either zero nodes or nearly the entire graph. Figure 5 shows a three-layer network of JOIN operations, creating an item AH that only fires when all eight

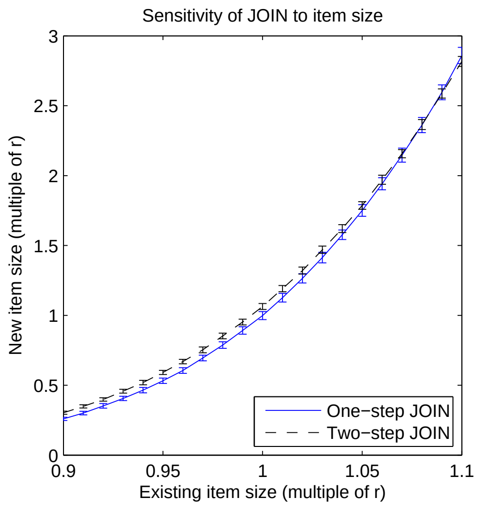


Fig. 4. The number of nodes in items created by the JOIN operation is extremely sensitive to variations in the number of nodes in the items being compounded. Using network parameters from [10], both one-step and two-step JOIN amplify small variations by approximately one order of magnitude.

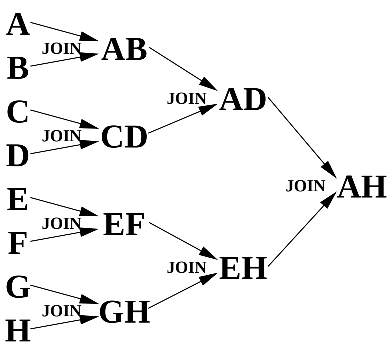
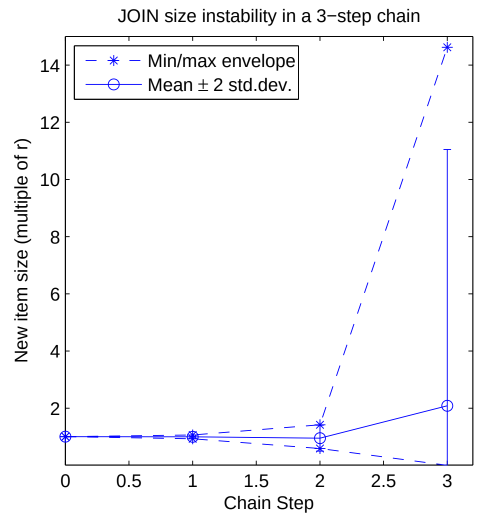
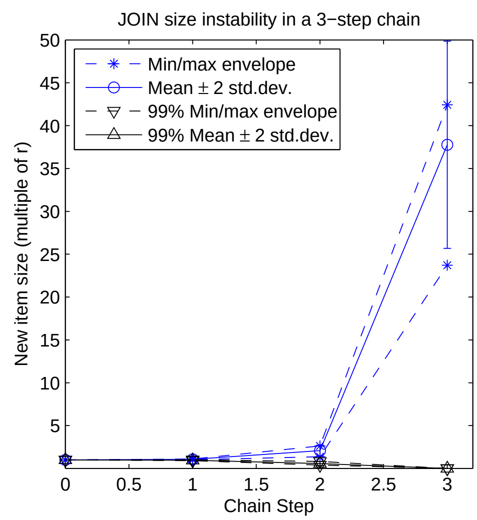


Fig. 5. The size sensitivity problem of JOIN is demonstrated by three layers of chained JOIN operations, compounding items A through H to form an item AH that fires only when all eight inputs are firing.



(a) One-step JOIN chain



(b) Two-step JOIN chain

Fig. 6. The size of items produced by JOIN operations explodes or collapses over the three-layer JOIN chain shown in Figure 5.

of the initial items A through H fire. We evaluate this with the same parameters as before, setting r for the initial items to the ideals $r = 2134$ for one-step JOIN, $r = 2338$ for two-step JOIN, gathering 100 data points for each (10 times on 10 networks). Results are shown in Figure 6. Since two-step JOIN always resulted in AH growing to cover the whole network, we also test two-step join with $r = 2314$—one percent less nodes in the initial items—and AH always collapses in size.

Thus, although the JOIN operation is viable in isolation, its sensitivity to the size of the items to be joined makes chaining JOIN operations impossible, at least as originally specified.

IV. Noise sensitivity in propagating signals

The composability of "circuits" in the sparse encoding model can be evaluated using the notion of *static discipline* from digital circuit design. Transfer curves are measured for JOIN and LINK circuits, determining how the fraction of output nodes firing varies with respect to the fraction of input nodes firing. These transfer curves determine the composability of signals: if appropriate noise margins can be chosen, then signals will be restored as they pass through circuits and noise poses no limit on composability. Otherwise, the circuits are sensitive to noise and signals can be expected to degrade, perhaps rapidly, as they pass through circuits.

In digital circuit design, devices are shown to be composable by establishing a static discipline—a relationship between input and output voltages that ensures that output voltages are closer to ideal "0" and "1" values than input voltages. More formally, the static discipline for a family of devices is a set of voltages levels, $V_{OL} < V_{IL} < V_{IH} < V_{OH}$. When device obeys the static discipline, we are guaranteed that if all input voltages V_i are below the low input threshold $V_i \leq V_{IL}$ or above the high input threshold $V_i \geq V_{IH}$, then the output of the device will be below V_{OL} if the output is a "0" and above V_{OH} if it is a "1." In other words, the standards for the digital values "0" and "1" are more stringent for outputs than for

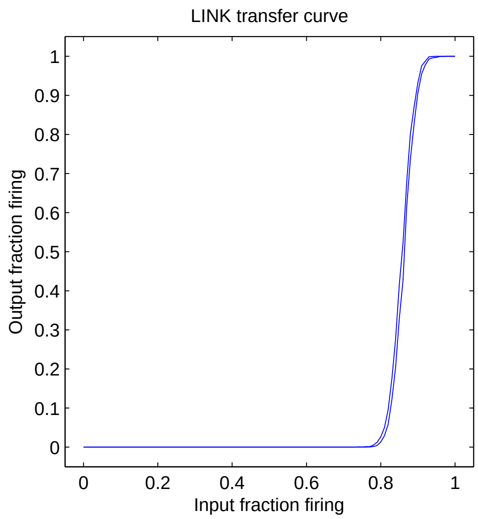
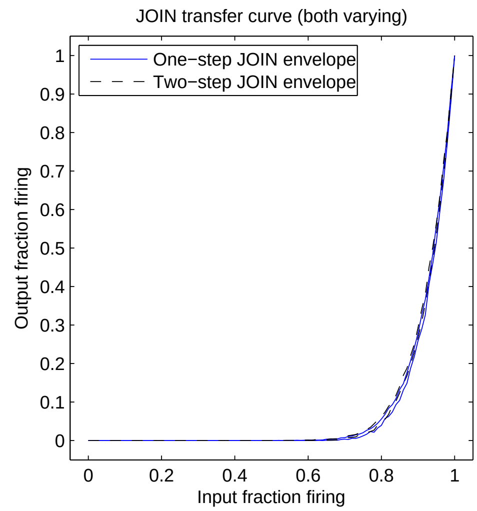
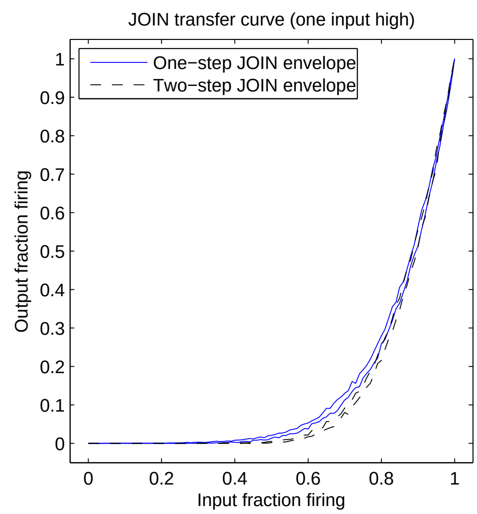


Fig. 7. Transfer curves showing the envelope of fraction of output nodes firing for 10 LINK circuits as the fraction of input nodes firing is varied. The flat top and bottom of the curve indicate that good noise margins can be established.



(a) JOIN, varying both inputs



(b) JOIN, one input held high

Fig. 8. Transfer curves showing the envelope of fraction of output nodes firing for 10 JOIN circuits as the fraction of input nodes firing is varied. No upper noise margin can be established, so even minimal noise will result in significant signal degradation.

inputs.

This restoration of voltages toward the ideal damps the effect of noise at each stage of a circuit, allowing digital devices to be composed without bound. The farther the input thresholds are from output thresholds, the greater the noise margin—the amount of noise that can be tolerated at each stage of the circuit.

The composability of JOIN and LINK circuits can be evaluated by a similar method, since items have the boolean values "recognized" and "not recognized," which may be interpreted as binary "1" and "0" respectively. The fraction f of an item's nodes that are firing plays the part of voltage, with no firing the ideal "0" and all firing the ideal "1."

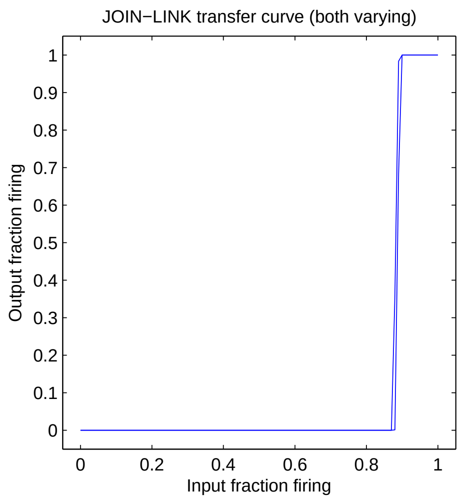
For each device, we measure its transfer curve by varying the fraction of inputs firing from 0 to 1 in steps of 0.01, using the parameters $n = 100,000$, $d = 512$, $r = 2134$, $k_a = 16$ and $k_m = 32$ for one-step JOIN, $k_m = 16$ for two-step JOIN (from [10]). For JOIN, we measure both the simultaneous variation of inputs and the variation of one input while the other is held high. Each transfer curve shows the envelope of behavior for 10 devices, each created on a different network. LINK circuits have good noise margins (Figure 7), but no upper noise margin can be established for JOIN (Figure 8): even minimal noise will result in significant signal degradation.

V. Amended Model

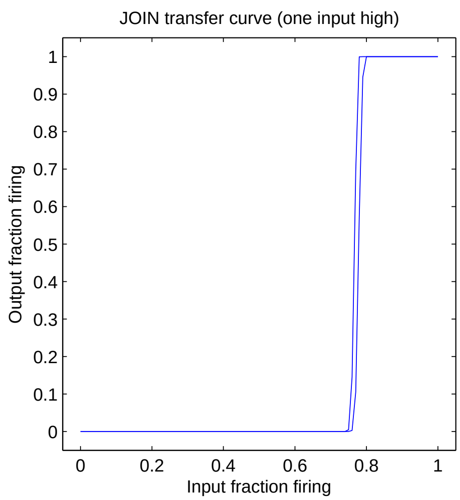
Although the original model of memorization is not viable under composition, two small modifications make it so. The first is to add an association stage to the end of a memorization circuit. This removes the size instability problem and steepens the slope of the transfer curve. The second is to lower the firing thresholds k_m and k_a slightly, shifting the transfer curve to provide an adequate noise threshold for firing items.

The new JOIN-LINK algorithm is simply a composition of the existing one-step JOIN and LINK algorithms. Given existing items A and B, first allocate a new random item C for the output. Use a one-step JOIN on A and B, creating an intermediate item γ, then LINK the item γ to C. The resulting circuit fires C if and only if both A and B are firing, as desired for memorization, though it takes three steps of propagating firings rather than one to execute. Because C is allocated independently from A and B, the JOIN-LINK algorithm does not suffer from instability in encoding size.

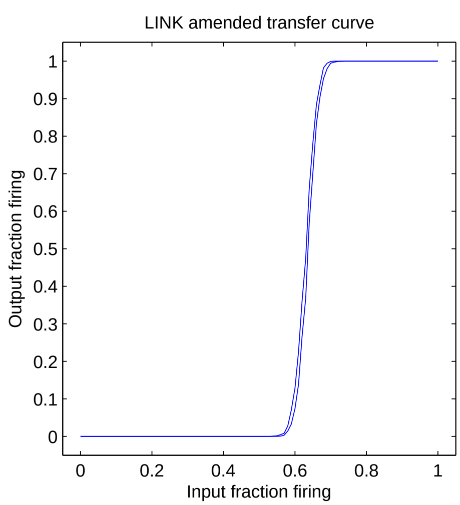
We measure transfer curves for JOIN-LINK and LINK as before, using the same $n = 100,000$, $r = 2134$, and $d = 512$, and lowered firing thresholds $k_m = 30$ and $k_a = 13$. The results (Figure 9) support a static discipline of $f_{IL} = 0.5$, $f_{OL} = 0.01$, $f_{IH} = 0.91$, and $f_{OH} = 0.99$, for a noise margin



(a) JOIN-LINK, varying both inputs



(b) JOIN-LINK, one input held high



(c) Amended LINK

Fig. 9. Measured transfer curve envelopes for the amended model, using $n = 100,000$, $r = 2134$, $d = 512$, $k_m = 30$ and $k_a = 13$, show composability, because a static discipline can be established with $f_{IL} = 0.5$, $f_{OL} = 0.01$, $f_{IH} = 0.91$, and $f_{OH} = 0.99$, for a noise margin of 8%.

of 8%. The amended model thus provides both stable encoding size and good noise margins, allowing unlimited composition with respect to construction and signal propagation.

VI. Contributions

Empirical evaluation shows that the memorization operations in the sparse encoding model proposed by Valiant are not composable: the size of items is sensitive to small variations, and signals propagating through memorization circuits degrade badly in the presence of even minimal noise. We therefore present an amended model, adding an association state to memorization but preserving the basic premises of the sparse encoding model. The amended model produces circuits that are not limited in their composition by the factors examined in this paper. We have thus filled in an important gap in the sparse encoding hypothesis.

Composition issues are not limited to the two we have addressed in this paper, and this work invites more investigation along similar lines. Particularly pressing are questions about the capacity of a sparse random graph to encode items and relationships and about the degree to which propagating signals interfere with one another. Preliminary investigation suggests that the original and amended model will both perform badly in these areas, but that once again a careful understanding of the problems will suggest small modifications to the model that correct this deficiency.

Ultimately, of course, the test for the sparse encoding model is comparison with actual mammalian cortex. However, there is not yet conclusive evidence to confirm or deny the two most fundamental assumptions of the model—sparse random graph structure and the importance of weak synapses. If the biological data does in the end support the sparse encoding model, the implications may be profound, for the digital computation in the model helps to explain how low-level neural activity might produce apparently "symbolic" higher-level cognition.

References

[1] M. Abeles. *Corticonics: Neural circuits of the cerebral cortex*. Cambridge University Press, 1991.
[2] A. B. Ali, J. Deuchars, H. Pawelzik, and A. M. Thomson. Ca1 pyramidal to basket and bistratified cell epsps: Dual intracellular recordings in rat hippocampal slices. *J. Physiology*, 507:201–217, 1998.
[3] R. L. Beurle. Properties of a mass of cells capable of regenerating impulses. *Philos. Trans. R. Soc. Lond. [Biol.]*, 240:55–87, 1955.
[4] V. Braitenberg and A. Schuz. *Cortex: Statistics and geometry of neuronal connectivity*. Springer-Verlag, 1998.
[5] J. Feldman and D. Ballard. Connectionist models and their properties. *Cognitive Science*, 6:205–254, 1982.
[6] J. S. Griffith. On the stability of brain-like structures. *Biophys. J.*, 3:299–308, 1963.
[7] H. Markram and M. Tsodyks. Redistribution of synaptic efficacy between pyramidal neurons. *Nature*, 382:807–810, 1996.
[8] L. Shastri. A computational model of episodic memory formation in the hippocampal system. *Neurocomputing*, pages 889–897, 2001.
[9] A. M. Thomson, J. Deuchars, and D. C. West. Large, deep layer pyramid-pyramid single axon epsps in slices of rat motor cortex displayed paired pulse and frequency-dependent depression, mediated presynaptically and self-facilitation, mediated postsynaptically. *J. Neurophysiology*, 70, 1993.
[10] L. Valiant. Memorization and association on a realistic neural model. *Neural Computation*, 17:527–555, 2005.