58th Annual IEEE Symposium on Foundations of Computer Science

Capacity of Neural Networks for Lifelong Learning of Composable Tasks

Leslie G. Valiant
John A. Paulson School of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02478, USA
valiant@seas.harvard.edu

***Abstract*—We investigate neural circuits in the exacting setting that (i) the acquisition of a piece of knowledge can occur from a single interaction, (ii) the result of each such interaction is a rapidly evaluatable subcircuit, (iii) hundreds of thousands of such subcircuits can be acquired in sequence without substantially degrading the earlier ones, and (iv) recall can be in the form of a rapid evaluation of a composition of subcircuits that have been so acquired at arbitrary different earlier times.**

We develop a complexity theory, in terms of asymptotically matching upper and lower bounds, on the capacity of a neural network for executing, in this setting, the following action, which we call *association*: Each action sets up a subcircuit so that the excitation of a chosen set of neurons A will in future cause the excitation of another chosen set B.

A succession of experiences, possibly over a lifetime, results in the realization of a complex set of subcircuits. The composability requirement constrains the model to ensure that, for each association as realized by a subcircuit, the excitation in the triggering set of neurons A is quantitatively similar to that in the triggered set B, and also that the unintended excitation in the rest of the system is negligible. These requirements ensure that chains of associations can be triggered

We first analyze what we call the Basic Mechanism, which uses only direct connections between neurons in the triggering set A and the target set B. We consider random networks of n neurons with expected number d of connections to and from each. We show that in the composable context capacity growth is limited by d^2, a severe limitation if the network is sparse, as it is in cortex. We go on to study the Expansive Mechanism, that additionally uses intermediate relay neurons which have high synaptic weights. For this mechanism we show that the capacity can grow as dn, to within logarithmic factors. From these two results it follows that in the composable regime, for the realistic cortical estimate of $d = n^{\frac{1}{2}}$, superlinear capacity of order $n^{\frac{3}{2}}$ in terms of the neuron numbers can be realized by the Expansive Mechanism, instead of the linear order n to which the Basic Mechanism is limited. More generally, for both mechanisms, we establish matching upper and lower bounds on capacity in terms of the parameters n, d, and the inverse maximum synaptic strength k.

The results as stated above assume that in a set of associations, a target B can be triggered by at most one set A. It can be shown that the capacities are similar if the number m of As that can trigger a B is greater than one but small, but become severely constrained if m exceeds a certain threshold.

***Keywords*-neuroscience; neural computation; neuroidal model; computational complexity; associations;**

I. Introduction

It is natural to seek an understanding of the computations cortex performs by considering what its fundamental capabilities are, and examining these in light of the resources that are available to it. For this pursuit one needs concrete formulations of the basic capabilities of cortex, and also a quantification of its most fundamental resources.

As far as capabilities we shall consider the tasks of memorization and learning. Besides formalizing what each act of memorization and learning accomplishes in isolation, as one would in learning theory, it is equally important to take into account the larger context that the brain is both (a) a self-contained system and (b) one that has to perform numerous tasks of multiple kinds. This context is somewhat unique in terms of computational systems that have been studied in a learning context. We shall therefore start by spelling out seven requirements that we regard as critical: (1) *One-Shot Memorization*: Some acts of knowledge acquisition require just one interaction with the outside world. (2) *Lifelong Learning*: In a lifetime hundreds of thousands of pieces of knowledge are acquired, and the functionality of each one is retained long after the time of its acquisition, and after much further knowledge has been acquired. (3) *Subcircuit Creation*: The acts of learning and memorization result in subcircuits that can be evaluated rapidly on new inputs, as in a recognition task, so as to realize the near instantaneous use of knowledge acquired over a possibly extended period. Such a subcircuit is not merely a representation of the knowledge but also an executable algorithm that acts out that knowledge. (4) *Composable Subcircuits*: A circuit acquired for a task should be of the form that its evaluation fits into a chain of evaluations of circuits learned at various earlier and later times. Knowledge acquired at disparate times can be combined on the fly, as needed, to perform a complex recognition task almost instantaneously. (5) *Quantitative Plausibility*: The basic resource requirements, of time steps, neuron numbers, connection numbers and synaptic strengths need to be plausible. (6) *Architectural Plausibility*: The algorithms should be executable on a distributed model that allows connections to have influence

DOI 10.1109/FOCS.2017.41

in just one direction, and needs minimal synchronization. (7) *Cognitively Adequate Set of Primitives*: A set of primitives must be supported simultaneously that might be a basis for implementing the broader functions of cognition. For this set we have previously advocated a set of four: association, supervised memorization, inductive learning of threshold functions, and hierarchical memorization [V94]. The intent is that these primitives be powerful enough in combination that a broad array of more complex cognitive functions be implementable in terms of them, as pursued, for example, in [PV15]. This paper focuses on the first of these primitives, association, but within a framework that has been shown to also support the three others.

The essential resources of cortex we shall capture as three numerical values: n, the number of neurons, d the number of connections from and to each neuron, and k, the inverse of the maximum strength that a connection, realized possibly as multiple synapses, from one neuron to another may have. For example, $k = 20$ means that when synapses are at their maximum permitted strength, 20 presynaptic neurons firing are enough to cause a postsynaptic neuron to fire. Each of these three resources takes up physical volume, and is therefore precious. A fourth resource is time: all our mechanisms are essentially optimal in this in requiring just one or two steps.

The results proved in this paper can be regarded as purely graph theoretic. They show that for a random directed graph where each edge has one of two weights, the requirements (1) - (6) above can be realized, with each learning act changing an appropriate set of weights, so that the graph when interpreted as a circuit of threshold elements realizes the composition of the functionality instances presented over a period. Humans appear to be able to realize perhaps hundreds of thousands of such functionality instances over lifetime. We are not aware of any proposals alternative to the one discussed in this paper that might account quantitatively for the scale of this phenomenon.

Of course, the human brain has far larger capacity in the pure information theoretic sense. But achieving substantial capacity while realizing the above computational functionalities with the architecture of the brain appears problematic.

As far as philosophical backdrop we note that while generalization is at the heart of the learning phenomenon, that is not what we study here. For the architecture of the brain a sharp resource tension arises already for the simpler task of association when implemented at scale.

II. The Neuroidal Model

The neuroidal model is a distributed model of computation designed to allow algorithms with the above described seven constraints to be described and analyzed [V94]. In this model the neurons and connections never change, only the synaptic strengths of the connections and the various states, as described below. The model is designed to *not overestimate* the capabilities of cortex: there should be little doubt that the algorithms that the model supports could be supported in cortex, and in a resource-faithful manner. It is designed as a framework in which mechanisms of quantitatively impressive functionality might be demonstrated to be supportable. Such a model is useful when, as in this paper, one can demonstrate mechanisms that are quantitatively consistent with what is known. Thesemwchanisms then offer an explanation of how cortex *might* compute. Currently there appears to be no alternative proposals that meet these constraints.

The model is entirely distributed, with each neuron running its own algorithm. The main features of the model are that each synapse i has a *weight* w_i, and the total memory capacity of a neuron includes not only these weights but also some additional *state* information that is distributed both at each synapse as a synapse state s_i and as an overall neuron state s. At any instant for any one neuron we denote the sum of weights w_i of all its synapses from presynaptic neurons that are firing, by w. The algorithms run on the neuron will update the states s, s_i and weights w_i under the following constraints: (i) a synaptic weight w_i or synaptic state s_i update can depend only on the weight w_i and state s_i of the same synapse, on the sum w, and on the value of the neuron state s, all at the previous instant, and (ii) the neuron state s can depend only on w and s at the previous instant. The model has a weak timing mechanism in which the neurons can stay synchronized for a few steps.

Regarding biological evidence we note that there appear to be a remarkable diversity of types among the neurons, and even among the synapses. From the neuroidal perspective, these reflect the diversity of algorithms that are implemented at different neurons.

Cortex is known to support global signaling by a small number of neuromodulators. However, since the total information content of these at any instant is negligible compared to that of the synapses, it is the distributed mechanisms realized at the neurons, and modeled by neuroids, that have to carry the main computational burden.

We note that in all our uses of the model we identify an $item$ (which may correspond to a real world concept, object or event) that is being processed, with a set A of neurons that are active. These sets of neurons, which we call $arbsets$, are not better connected to each other than arbitrary sets, and are in that way unlike Hebb's $assemblies$, which are.

III. Relations to Previous Work

There is a large literature on "association" neural networks, in various senses of that word [WBLH69, W71, H82, BW92, GW97, BK98, DA03, KPS06, ELS14]. Our emphasis on composability points to "hetero-associative" networks in the style of Willshaw [WBLH69, W71, GW97] as opposed to "auto-associative" networks. The basic Willshaw network, as reformulated for the present context, is as follows: There is a network consisting of n input nodes and n output nodes, with directed edges going from the former to the latter. For randomly chosen sets $X_1, ..., X_C$ of size R of input nodes and randomly chosen sets $Y_1, ..., Y_C$ of size r of output nodes, set up the C associations so that in the future: whenever all the neurons in one X_i (and no other neurons) fire then the result will be that all of the corresponding Y_i will fire, but with significant probability none of the other neurons will.

For computations in the brain we need to model sparse networks, where the expected number of connections from and to each neuron is $d \ll n$. We are particularly interested in the case $d = n^{\frac{1}{2}}$, which we regard as illustrative of mammalian cortex. If the synapses are binary and there are $O(nd)$ modifiable synapses, then $2^{O(nd)}$ different combinations of weights can be realized. Then it follows that the capacity C can be at most $\tilde{O}(nd)$, where the $\tilde{O}$ notation means that logarithmic factors are suppressed. The reason is that if the sets $X_1, ..., X_C, Y_1, ..., Y_C$ are all different and fixed, then $C!$ different permutations of them need to be realizable. But $C!$ is less than $2^{C \log_2 C} = 2^{\tilde{\Omega}(C)}$, and hence for this not to exceed $2^{O(nd)}$, C can be at most $\tilde{O}(nd)$.

We note that the Willshaw formulation is well suited to the property of composability in one important respect: When X_i fire, very little outside Y_i will fire, so that if Y_i were an input to a second association then chaining can proceed without substantial extraneous activity developing. To allow for composability we shall in our formulation make two modifications: First we allow the X_i, Y_i to be subsets of a common set, rather than from disjoint input and output sets, so as to permit chaining of associations. Second, we insist that X_i, Y_i are all of the same size, which, as we shall see, is not where capacity optimizes for Willshaw nets.

We make the following additional observations about how our results differ from earlier analyses of Willshaw nets. First, our formulation requires only local knowledge to determine whether a neuron fires. Some earlier definitions assume a winner-take-all criterion, where the threshold for firing is defined in terms of the relative activity levels at all the neurons. Second, some earlier analyses seek to optimize the amount of information in the triggered sets of neurons. In our analysis the number of bits of information contained in the description of a Y_i is not relevant.

Using an informal argument Graham and Willshaw [GW97, Appendix B] argue that an order of growth of $C = \tilde{\Theta}(nd)$ is indeed achievable for their formulation. However, to achieve it they need that the ratio $\frac{R}{r}$, where R is the size of the input sets X_i and r the size of the output sets Y_i, grow as $\frac{n}{d}$. For the parameters of relevance, particularly $d = n^{\frac{1}{2}}$, this means that the sizes R, r of the inputs and outputs are greatly unbalanced. This is at odds with the need that the circuits be composable, which requires that $r = R$, at least approximately, and in which case their capacity estimate becomes $\tilde{\Theta}(d^2)$. Note that these capacity estimates of Graham and Willshaw [GW97] are consistent with our upper and lower bounds for the Basic Mechanism (Theorems 1 and 2).

With regard to other neural models, we note that for networks that perform auto-association, that is those that learn a set of bit-strings and can recover the closest one given a noisy version [H82], capacity $\Theta(nd)$ can indeed be achieved [DA01]. But this auto-associative functionality has not been shown to be able to support composable circuits with a cognitively adequate set of primitives in the sense discussed here. In a different direction, we note that our relay nodes differ from conventional "hidden units" in requiring no updating, and hence evade that objection to biological plausibility.

IV. Random Graphs and Random Sets of Neurons

We consider random directed graphs $G_{n,d} = (V, E)$ on n neurons V where the edges E consist of ordered pairs (x, y) of neurons joined by a directed edge, in each of the two directions independently with probability d/n. (We permit self-loops, but do not exploit them.) For a graph G and sets $X_1, ..., X_C, Y_1, ...Y_C \subset V$ we will seek to realize associations $X_i \to Y_i$ for $1 \leq i \leq C$ in the sense of the section to follow. Each X_i will have R nodes and each Y_i r nodes. We particularly have in mind the *composable setting*, where $R = r$ and some of the triggered sets Y_j are equal to some triggering set X_i, so that chains of associations, such as $X_1 \to Y_1 = X_2 \to Y_2$ can be realized by a single circuit evaluation. We think of the Y_i as randomly generated first for $i = 1, ..., C$. Then each X_i is either randomly generated independent of everything prior, or is made equal to some Y_j with $j < i$, but no Y_j is equated with more than one X_i.

V. Associations

For simplicity, and as will be sufficient, we shall consider here that synaptic strengths can have one of just two possible values, zero and a *high* value. We take the latter to be $\frac{1}{k}$ if we set the thresholds of all the neurons to be unity, so that a neuron will fire whenever at least k of its high valued synapses come from neurons that are firing.

We say that neuron set X *directly t-excites* neuron y if the number of high valued edges incoming to y from X is at least t. We note that if $y \in X$ then whenever X all fire so will y, but we will say that X t-excites y only if such an influence is realized entirely via edges that leave X and arrive back at y. In general, X_i and Y_i are from a common set of neurons, so that some X_i may intersect with some Y_j and indeed may be the same.

In the following definitions of capacity probabilities arise both from the selection of the Xs and Ys, and also from the selection of the edges E of G. By a *mechanism* we mean a neuroidal algorithm. In Sections 6 and 8 we shall describe particular such mechanisms. For general such mechanisms we say that neuron set X *t-excites* neuron y if when all of

X fire, but no others, then y will be directly t-excited by a set of neurons from which it receives connections.

Definition 1 We say that a mechanism *has capacity C for* (k, k', R, r)*-associations* for some $k' \leq k$ if with respect to randomly chosen sets $X_1, ..., X_C \subset V$ of size R and $Y_1, ..., Y_C \subset V$ of size r the mechanism achieves that: (A) for each i and for each $y \in Y_i$ the probability is more than $1 - \frac{1}{2rC}$ that X_i k-excites y, and (B) for each i the expectation of the number of neurons y such that $y \notin Y_i$ and X_i k'-excites y, is less than $\frac{1}{2}$.

We now define the corresponding notion in the composable setting where $r = R$ and chaining is possible.

Definition 2 We say that a mechanism *has capacity C for* (k, k', r)*-composable-associations* for some $k' \leq k$ if for randomly chosen sets $Y_1, ..., Y_C \subset V$ of size r, and sets $X_1, ..., X_C$ also of size r, each chosen either randomly or to be equal to some Y_i with no two Xs being the same Y_i the mechanism achieves that: (A) for each i and for each $y \in Y_i$ the probability is more than $1 - \frac{1}{2rC}$ that X_i k-excites y, and (B) for each i the expectation of the number of neurons y such that $y \notin Y_i$ and X_i k'-excites y, is less than $\frac{1}{2}$.

The intent of (A) is to ensure that there is at least an even chance that all of the C associations $X_i \to Y_i$ are implemented exactly: If there are C choices of i, and, for each i there are r choices of y, then a maximum $\frac{1}{2rC}$ probability of failure for each combination, will, by the union bound, imply an upper bound of $\frac{1}{2}$ on the probability of any failures among those rC events.

The intent of (B) is to ensure, roughly, that in expectation, for each X_i the odds are against there being any $y \notin Y_i$ that is k'-excited by X_i.

We note that the default case is that $k' = k$. Smaller k' is harder to achieve but potentially offers more resilience to noise. For our lower bound results we choose k' substantially less than k, such as $\frac{k}{8}$, simply to make the proofs easier.

These definitions assume that in a set of associations, a target B can be triggered by at most one set A. It can be shown that the capacities are similar if the number m of As that can trigger a B is greater than one but small, but become severely constrained if m exceeds a certain threshold.

VI. The Basic Mechanism

We first explore the capabilities of a particular simple update algorithm, which we shall call the Basic Mechanism. We shall later go on to consider the slightly more complex Expansive Mechanism that builds on it. We shall prove bounds on the capabilities and limitations of both, with a view to demonstrating the superior capacity of the latter in the composable context.

In the Basic Mechanism each neuron has threshold 1, and initially each synapse has the low value of 0. When an association $X_i \to Y_i$ is being learned each neuron $y \in Y_i$ goes into a certain state s and all the neurons in X_i are made to fire. In state s a neuron executes the following update algorithm: *For edges incoming from nonfiring neurons the weights are left unchanged. For edges incoming from firing neurons the weight of each is set to* $1/k$.

The result of an execution of this mechanism at a neuron $y \in Y_i$ is the following. If $A \subset X_i$ is the set of its presynaptic neighbors that fire during the execution of this mechanism and $|A| \geq k$, then at subsequent times whenever all the neurons in X_i fire, so will the target neuron y since at least k presynaptic neurons will be firing via edges each with weight $\frac{1}{k}$. If $|A| < k$ then all the edges from A to y will be also set to $\frac{1}{k}$, but the subsequent firing of A will not cause y to fire because the threshold would not have been reached. Further, only synapses on edges from the X_i to the Y_i neurons will have changed.

This algorithm can be easily implemented in the neuroidal model. When realizing $X_i \to Y_i$ both the sets X_i and Y_i will be caused to fire at some point, as a result perhaps of some input acting through some other subcircuits. This will force the set Y_i to go into the special state s from which it will execute the described mechanism, while other neurons, that are not in that state, will not.

The Basic Mechanism may be viewed as the most natural mechanism for networks in which there are direct connections from the X_i to the Y_i. It is equivalent to that used in Willshaw nets [W71].

When the Basic Mechanism has been realized for a set of C associations on graph $G = (V, E)$, we say that the *effectuating edge set* is $E^* = E \cap (\bigcup_i \{(x, y) | x \in X_i, y \in Y_i\})$. This is just the set of edges that have been assigned high values by the mechanism.

VII. Relations Among n, d, r, R, k for the Basic Mechanism

We have five positive integer parameters n, R, r, k, d. It is clear by definition that $d \leq n$ since $\frac{d}{n}$ is a probability, that of a connection from one neuron to another. Also, if we have even a single source item X, represented by R neurons, that can cause another neuron y to fire that has indegree d (the average), then for any association to be realized $k \leq R$ and $k \leq d$, since y requires at least k inputs to come from X. Also, for nontriviality we assume $r, R < n$.

We now turn to the capacity C, the number of associations that can be acquired in succession without the ones acquired earlier degrading in their effectiveness. A basic estimate of the resources needed for these operations can be derived from the observation that there are r neurons in a target item Y, and each one needs at least k incoming synapses. Hence each component $X_i \to Y_i$ of an association corresponds to at least kr synapses. Now, there is an expected number of dn synapses in the whole system. Hence a *default estimate* is that the capacity of the system is upper bounded by $\nu = \frac{dn}{kr}$. (This corresponds to the maximum possible count obtained when the synapses are not shared among the different components of a set of associations.) We shall show that for both the Basic Mechanism and the Expansive Mechanisms this default estimate is indeed a provable upper bound to a constant factor (Theorems 2(b) and 4(b)). We shall also show that it can be achieved to constant factors (Theorems 1(b) and 3(b)) for relevant parameters. Thus, while the mechanisms do allow synapse sharing, the capacity they achieve cannot exceed the default estimate.

A further ratio that will emerge from the analysis is the *basic efficacy* $\eta_b = \frac{kn}{Rd}$. We shall show (Theorem 2(a)) that the Basic Mechanism imposes the constraint $\eta_b < 1$. For the Expansive Mechanism the same holds for the *expansive efficacy* $\eta_e = \frac{kn}{RDd}$ (Theorem 4(a)).

In [V06], where k was defined as the inverse mean rather than inverse maximum synaptic strength, it was pointed out that if for that k, $\frac{kn}{Rd} \geq 1$ then the firing of a single neuron will cause too many others to fire. For the Standard Mechanism in this paper, for example, this inverse mean will grow from zero slowly as associations are accumulated, and the capacity will be exceeded before it gets near the inverse maximum.

VIII. The Expansive Mechanism

The Expansive Mechanism is like the Basic Mechanism except that it is realized via an intermediate layer of *relay* neurons which aid in the distribution of information without increasing the degree. The inputs X_i and outputs Y_i of the associations are represented in a *basis* layer of n neurons. The basis neurons are connected in a forward direction to a set of neurons in the relay layer, which, for simplicity, we also assume to number n. A forward connection between an arbitrary neuron in the basis layer and an arbitrary neuron in the relay layer exists with probability $\frac{D}{n}$, so that the expected number of connections from a basis neuron or to a relay neuron is D. The synaptic weight on an incoming edge to a relay neuron is fixed to be 1. Thus when the R neurons in X_i fire then a possibly much larger set A_i of up to about RD relay nodes will be caused to fire, and about this number will fire if $RD \ll n$.

There is a second set of edges directed from the relay neurons to the basis neurons. The probability of a connection from a fixed relay neuron to a fixed basis neuron is d/n, so that the expected number of connections to a basis neuron or from a relay neuron is d. A random graph with basis and relay neurons as described and having these parameters, we call $G_{n,d,D}$.

Philosophically, either of the sets X_i or A_i (or both) in the above discussion can be regarded as "representing" the item at hand. (Thus if X_i is small and A_i is large then this is both a sparse and a dense representation!) Having a large A_i gives the benefit of wider connectivity into the whole system. However, the only synapses that are modifiable are those from the relay to the basis layer, which initially have the value 0. *For realizing the association $X_i \to Y_i$ the Expansive Mechanism updates the synapses from the set A_i of relays to the neuron $y \in Y_i$ in the basis layer exactly as the Basic Mechanism acts between an input set X_i and the neuron y.*

The *effectuating set* of edges will be defined here as those edges from the relay to the basis layer that are set high by the mechanism for the set of C associations at hand. Taking the view of the default estimate of C as $\frac{nd}{rk}$ or, equivalently, that synapses are not substantially reusable for different associations, a smaller r ensures that fewer synapses are devoted to any one item, and hence larger capacities are possible.

In the Basic Mechanism the same edges have to realize both the representation (where small may be good) as well as the connectivity (where large may be good.) In the Expansive Mechanism these two tasks are in a certain sense split. Most notably, in order to achieve high capacity the Basic Mechanism needs $|X_i| = R$ much larger than $|Y_i| = r$. One can hypothesize that the Expansive Mechanism simulates such a larger R using $|A_i|$ as proxy. As the analysis in the proof of Theorem 3 shows, this hypothesis turns out to be true. But this needs proof since, unlike the X_i, the A_i are by no means independent of each other for the different i.

IX. Relations Among n, d, r, R, k, D for the Expansive Mechanism

As far as our new parameter D it is clear, by definition, that $D \leq n$ since $\frac{D}{n}$ is the probability of a connection. As in the Basic Mechanism the *default estimate* of the capacity is again $\nu = \frac{nd}{kr}$, since the weights of the relay neurons are never changed, and among the basis neurons, about kr synapses are associated with any one association out of a total number of dn synapses. However, we now define the *expansive efficacy* to be $\eta_e = \frac{kn}{RDd}$. We will find in Theorems 3 and 4 that this quantity plays a similar role as η_b, and that $\eta_e < 1$ holds in this new context also.

X. Summary of Our Results

The dependence of the capacity C on all the other parameters is quite complex. The following summarizes some of the more relevant dependencies shown by of our results. *Some are proved under certain technical constraints on the parameters.*

We start with the general case where r, R are not necessarily equal, and give a general upper bound (Theorem 2) on capacity:

(I) *For the Basic Mechanism $C \leq \frac{d^2 R}{k^2 r}$.*

In the composable setting, where $R = r$, it follows that d^2 is an upper bound on capacity, which is linear in n for the density $d = n^{\frac{1}{2}}$ that is of most interest, and falls short of the information theoretically optimal capacity of $\Theta(dn)$.

On the other hand, we show that the optimal capacity dn is achievable for suitable parameters, but this needs R to be larger than r (Corollary 1.1):

(II) *For every $c > 0$ there exist $\lambda, K > 0$ such that capacity $C = \lambda \frac{dn}{(\log_2 n)^2}$ can be achieved by the Basic Mechanism with $k = K \log_2 n$, $r = 3k$, $R = \frac{nr}{d}$ and $k' = ck$.*

Thus, unless the density d is close to n, this capacity $C = \tilde{\Theta}(nd)$ is achieved with disparate R, r sizes, which is inconsistent with composability.

We go on to show that with the slightly more complex Expansive Mechanism, optimal capacity $C = \tilde{\Theta}(nd)$ can be realized with $R = r$, as required for composability (Corollary 3.1 to Theorem 3):

(III) *For any constant $c > 0$, for $d = D = n^{\frac{1}{2}}, r = R = 3k$, and $k = Klog_2 n$ for large enough constant K, capacity $C = \Theta(\frac{n^{\frac{3}{2}}}{(\log_2 n)^2})$ can be achieved by the Expansive Mechanism.*

Here d is still the expected degree of the network, but now there is an added layer of *relay* neurons, and all paths from the X_i to the Y_i through which excitation occurs are of length two and go through relay neurons. Our present result is a quantitative justification of both (a) the usefulness of relay node algorithms as introduced in [V94], as well as (b) the *the Strong Synapse Hypothesis* [V94, Chap 14.2] that posits that for tasks such as associations the existence of some strong synapses are advantageous. It confirms some previous evidence from simulations [FV09].

Theorem 4 gives a limitation of the following kind:

(IV) *The claims of Theorem 3 are optimal to constant factors for the Expansive Mechanism, within wide ranges of the parameter values.*

One can also consider *many-to-one associations*, where the same Y is the target of m distinct Xs. This generalization is necessary to reflect the fact that in cognition several concepts X may need to be associated to the same target Y. The following is an informal statement of what can be proved:

(V) *For both the Basic and the Expansive Mechanism the upper bounds stated for $m = 1$ hold also for moderate values of $m > 1$. However, for $m > \frac{n}{R}$ the capacity C drops precipitously.*

All our definitions and results are formulated so as to make the proofs as transparent as possible. The constant multipliers provided are mostly only for illustrative purposes and can be improved. The results have numerous technical constraints, retained in unoptimized form so that their role can be easily traced.

XI. Some Combinatorial Inequalities

We shall use several well-known properties of the binomial distribution. Let $S_1, ..., S_m$ be a sequence of m independent $\{0, 1\}$-valued random variables where, for all i, $\Pr[S_i = 1] = p$. Let S be the sum of the S_i, and let $\mu = mp$ be the expectation of S. Let $\beta^{>}(m, \mu, \alpha)$ be the probability $\Pr[S > \alpha\mu]$ and let $\beta^{\geq}(m, \mu, \alpha)$ be the probability $\Pr[S \geq \alpha\mu]$. Sometimes it is convenient to use the threshold $\tau = \alpha\mu$ (rather than α) as the third parameter, in which case we denote $\beta^{>}(m, \mu, \alpha)$ by $\gamma^{>}(m, \mu, \tau)$ and $\beta^{\geq}(m, \mu, \alpha)$ by $\gamma^{\geq}(m, \mu, \tau)$.

For bounding such tails of the Binomial distribution we shall use Chernoff bounds. Some Chernoff bounds also hold [PS97] when the m random variables are not independent, but only *negatively correlated* in the following senses: The set of $\{0, 1\}$-valued random variables with $\Pr[S_i = 1] = p$ is *negatively upper correlated* if for any subset $B \subset \{1, ..., m\}$ the probability that $S_i = 1$ for all $i \in B$ is at *most* $p^{|B|}$. The set is *negatively lower correlated* if for any subset $B \subset \{1, ..., m\}$ the probability that $S_i = 0$ for all $i \in B$ is at *most* $(1 - p)^{|B|}$.

Proposition 1 (Chernoff bounds)
(i) for $0 < \alpha \leq 1$, $\beta^{\geq}(m, \mu, \alpha) \geq 1 - (e^{\alpha-1}/(\alpha^{\alpha}))^{\mu}$.
(ii) for $\alpha \geq 1$, $\beta^{\geq}(m, \mu, \alpha) \leq (e^{\alpha-1}/(\alpha^{\alpha}))^{\mu}$.
(iii) for $0 < \alpha \leq 1$, $\beta^{>}(m, \mu, \alpha) > 1 - (e^{\alpha-1}/(\alpha^{\alpha}))^{\mu}$.
(iv) for $\alpha \geq 1$, $\beta^{>}(m, \mu, \alpha) < (e^{\alpha-1}/(\alpha^{\alpha}))^{\mu}$.
(v) The bounds (i) and (iii) hold for negatively lower correlated sets of random variables with the same m, p, α. The bounds (ii) and (iv) hold for negatively upper correlated sets of random variables with the same m, p, α.

Proposition 2 *For the binomial distribution for m independent identically distributed events with expectation μ:*
(i) the median is either $\lceil \mu \rceil$ or $\lfloor \mu \rfloor$, or both if μ is an integer.
(ii) if $m > 2$ then Pr$[S \geq \lceil \mu \rceil] \leq \frac{3}{4}$.

We shall also use the following, where part (i) is Bernoulli's inequality, and part (ii) follows from the fact that $1 + x \leq e^x$ for all real values of x.

Proposition 3 *For all $0 \leq x \leq 1$,*
(i) if $y \geq 1$ then $(1 - x)^y \geq 1 - xy$, and
(ii) if $y > 0$ and $xy \leq 1$ then $(1 - x)^y < 1 - \frac{1}{2}xy$.

Proposition 4 *Suppose $H_{i,j}$ for $1 \leq i \leq n$ and $1 \leq j \leq m$ are $\{0, 1\}$-valued random variables such that for each i the set $\{H_{i,j} | 1 \leq j \leq m\}$ is negatively lower correlated, and that the only correlations are among variables with the same i. Then the set $\{\bigwedge_{1 \leq j \leq m} H_{i,j} | 1 \leq i \leq n\}$ is also negatively lower correlated. The analogous proposition also holds for negatively upper correlated sets.*

Proposition 5(i) *For two randomly and independently chosen sets $A, B \subset T$ where $|T| = n$ the probability that $|A \cap B| \geq 4\frac{|A||B|}{n}$ is at most $\gamma^{\geq}(|A|, 2\frac{|A||B|}{n}, 4\frac{|A||B|}{n}) \leq (\frac{e}{4})^{\frac{2|A||B|}{n}}$.*

Proof First we note that if $|A| > \frac{n}{4}$ then $4\frac{|A||B|}{n} > |B|$, and $|A \cap B| \le 4\frac{|A||B|}{n}$ with certainty. Otherwise, we can consider a process in which the members of B are fixed and the members of A are chosen in succession at random, independent of previous draws. If m members of B have been found so far among the choices for A and $|A| \le \frac{n}{4}$, then the probability of hitting another member of B is $\frac{|B|-m}{n-|A|} \le \frac{4|B|}{3n}$. Hence the probability of finding another element in B is no greater than $\frac{2|B|}{n}$ each time. Applying Proposition 1(ii) gives the second bound.

Proposition 5(ii) *For two randomly chosen sets $A, B \subset T$ where $|T| = n$ and $|B| \ge 2|A|$, the probability that $|A \cap B| \ge \frac{|A||B|}{4n}$ is at least $1 - \gamma^{\ge}(|A|, \frac{|A||B|}{2n}, \frac{|A||B|}{4n}) > 1 - (\frac{2}{e})^{\frac{|A||B|}{4n}}$.*

Proof We can consider a process in which the members of A are chosen in succession with a probability of intersecting B at least $\frac{|B|}{2n}$ each time. To see this note that after m members of A have been chosen, with $m \le |A| \le \frac{|B|}{2}$, there are still $|B| - m \ge \frac{|B|}{2}$ candidates left, and and at most n to choose from. Hence the probability of success is at least $\frac{|B|}{2n}$ at each stage. Applying Proposition 1(i) gives the second bound.

XII. Realizing Associations by the Basic Mechanism: Possibilities

We shall first show that the Basic Mechanism can indeed achieve substantial capacity. In particular, Corollary 1.1 will assert that, for certain parameters, capacity $C = \tilde{\Theta}(nd)$ is achievable, and Corollary 1.2 that under the composability constraint $r = R$, the more limited capacity $C = \tilde{\Theta}(d^2)$ is achievable. Note that Theorem 2 will show that the assumption $\eta_b < 1$ in Theorem 1 holds unconditionally.

Theorem 1 *For all positive integers $C > 1, n, d, k, R, r$ and k', where also $c = \frac{k'}{k}, \nu = \frac{nd}{kr}$ and $\eta_b = \frac{kn}{Rd} \le 1$, if (a) $C < \frac{1}{2r} e^{k(\frac{1}{\eta_b}-1)} \eta_b^k$ and (b) $C \le \frac{ckn^3}{4erR^2d} = \frac{c}{4e}\eta_b^2\nu$ then on random graph $G_{n,d}$ the Basic Mechanism realizes (k, k', R, r)-associations to capacity C, provided that also (i) $(\frac{e}{4})^{\frac{Cr}{n}} < \frac{1}{4n}$ and (ii) $2 + \log_2 n < ck$.*

Proof (A) We first show that (a) is enough to guarantee that part (A) of Definition 1, of (k, k', R, r)-associations, is satisfied. Consider some i with $1 \le i \le C$ and the associated randomly chosen sets X_i, Y_i. Fix i, X_i and Y_i, and consider G as randomly generated from $G_{n,d}$. Then for each $y \in Y_i$ the probability that at least k edges incoming to y are set to high by the association $X_i \to Y_i$ is $\gamma^{\ge}(R, \frac{Rd}{n}, k) = \beta^{\ge}(R, \frac{Rd}{n}, \frac{kn}{Rd})$, the probability in R independent Bernoulli trials (corresponding to the R neurons $x \in X_i$), each with probability $\frac{d}{n}$ of success (that (x, y) is in G) that there are at least k successes. To satisfy condition (A) it is sufficient that $\beta^{\ge}(R, \frac{Rd}{n}, \frac{kn}{Rd}) \ge 1 - \frac{1}{2rC}$. For this it is sufficient that k be sufficiently smaller than the expectation Rd/n. Also, for $\alpha = \frac{kn}{Rd} = \eta_b$ it is the case that $\alpha = \eta_b < 1$ by assumption. Hence we can apply Proposition 1(i) with $\mu = Rd/n$ and $\alpha = \frac{kn}{Rd}$ to get $\beta^{\ge}(R, \frac{Rd}{n}, \frac{kn}{Rd}) \ge 1 - e^{-\mu}(e/\alpha)^{\alpha\mu} = 1 - e^{-\frac{Rd}{n}}(\frac{edR}{kn})^k$. Hence for this $\beta^{\ge}(R, \frac{Rd}{n}, \frac{kn}{Rd})$ to be greater than $1 - \frac{1}{2rC}$, it is sufficient that $2rC < e^{Rd/n}(\frac{edR}{kn})^{-k}$, or that $C < \frac{1}{2r}e^{\frac{Rd}{n}-k}(\frac{kn}{Rd})^k$.

(B) We now show that part (B) of the definition of (k, k', R, r)-associations is also satisfied by the conditions of this Theorem. Consider $G = (V, E)$ as randomly generated from $G_{n,d}$, and let $y \in V$. Let I_y be the set of neurons that are connected to y by edges in G, and let Q_y be the associations s, $1 \le s \le C$, such that $y \in Y_s$.

Now for any one y there are C choices of s, each having probability $\frac{r}{n}$ of success, namely of $y \in Y_s$. Hence the probability of more than $\frac{2Cr}{n}$ successes is $\gamma^{>}(C, \frac{Cr}{n}, \frac{2Cr}{n}) = \beta^{>}(C, \frac{Cr}{n}, 2)$, which by Proposition 1(iv) is less than $(\frac{e}{4})^{\frac{Cr}{n}}$. Hence by Condition (i), and summing expectations for all the y, the expected number of neurons y with $|Q_y| > \frac{2Cr}{n}$ is less than $\frac{1}{4}$.

Consider now that G has been chosen. Fix an i, Y_i and a $y \notin Y_i$ and assume that $|Q_y| \le \frac{2Cr}{n}$. Consider X_i as randomly generated. The probability for some particular $x \in V$ that "$x \in X_i$ and (x, y) is in the effectuated set" is $p_{x,y} \le \frac{R}{n}\frac{d}{n}(1 - (1 - \frac{R}{n})^{\frac{2rC}{n}})$, since $\frac{R}{n}$ is the probabilty that $x \in X_i$, $\frac{d}{n}$ is the probability that $(x, y) \in E$, $(1 - (1 - \frac{R}{n})^{|Q_y|})$ is the probability that $x \in X_s$ for at least one $s \in Q_y$, and $|Q_y| \le \frac{2rC}{n}$ has been assumed. This $p_{x,y}$, by Proposition 3(i), is at most $\frac{2CrdR^2}{n^4}$. (Note that this Proposition applies since $\frac{R}{n} < 1$ and $\frac{2rC}{n} \ge 1$ by virtue of (i).)

Now each of the three multiplicands in the expression for $p_{x,y}$ represents negatively upper correlated events for any fixed y as x varies. First, the events that $x \in X_i$ for the various x, are negatively upper correlated since for any $S \subset V$, the probability Prob$(\bigwedge_{x \in S} x \in X_i) \le (\frac{R}{n})^{|S|}$. This follows inductively from the fact that Prob$(v_j \in X_i | \bigwedge_{l<j} v_l \in X_i) = \frac{R-j}{n-j} \le \frac{R}{n}$ where $S = \{v_1, ..., v_{|S|}\}$ Second the events that $(x, y) \in E$ are independent for the different x. Finally, the events that $x \in X_s$ for at least one $s \in Q_y$ can be seen to be negatively upper correlated also for the different x.

We conclude that, by virtue of Proposition 4, for the fixed y the events that (x, y) are in the effectuated set for the different x are negatively upper correlated. In Proposition 4 for the H_{ij} the i correspond to the xs and $j \in \{1, 2, 3\}$ correspond to the three multiplicands in $p_{x,y}$. Hence by Proposition 1(v) we can use the same Chernoff bounds as hold if the events were independent.

Applying Proposition 1(ii), the probability that at least t of these edges (x, y) have "$x \in X_i$ and are in the effectuated set" is at most $\gamma^{\ge}(n, np_{x,y}, t) \le \gamma^{\ge}(n, \frac{2CrdR^2}{n^3}, t) \le (\frac{2dCerR^2}{n^3 t})^t$ provided the condition Z that $t \ge \frac{2dCrR^2}{n^3}$ holds. Hence, under this condition Z, the probability that X_i

directly $t-$excites y is upper bounded by $p_y = (\frac{2dCerR^2}{n^3 t})^t$, which is at most $(\frac{1}{2})^t$ if $C \leq \frac{ckn^3}{4erR^2d}$ and $t = ck$.

But the needed condition Z is satisfied since putting $t = ck$ and $C \leq \frac{ckn^3}{4erR^2d}$ guarantees that $t = ck \geq \frac{2dCrR^2}{n^3}$.

We deduce that the probability p_y is upper bounded by $(\frac{1}{2})^{ck}$. If this is less than $\frac{1}{4n}$, as is guaranteed by condition (ii), then the expected number of neurons $y \notin Y_i$ with $|Q_y| \leq \frac{2Cr}{n}$, that are ck-excited by X_i is less than $\frac{1}{4}$.

Hence we deduce, using the union bound, that the expected number of $y \notin Y_i$ that are ck-excited by X_i, whether satisfying $|Q_y| \leq \frac{2Cr}{n}$ or not, is less than $\frac{1}{2}$ as is required by condition (B).

Corollary 1.1 *For every $c > 0$ there exist $\lambda, K > 0$ such that capacity $C = \lambda \frac{dn}{(\log_2 n)^2}$ can be achieved with $k = K \log_2 n$, $r = 3k$, $R = \frac{nr}{d}$ and $k' = ck$, provided (i) $(\frac{e}{4})^{\frac{Cr}{n}} < \frac{1}{4n}$ and (ii) $2 + \log_2 n < ck$.*

Proof By substitution into parts (a) and (b) of the Theorem. Part (a) can be satisfied for any λ by a suitable K, and part (b) can be satisfied for any c, k for a suitable λ.

Corollary 1.2 *For every $c > 0$ there exist $\lambda, K > 0$ such that capacity $C = \lambda (\frac{d}{k})^2$ can be achieved with $k = K \log_2 n$, $R = r = \frac{3kn}{d}$ and $k' = ck$, provided (i) $(\frac{e}{4})^{\frac{Cr}{n}} < \frac{1}{4n}$ and (ii) $2 + \log_2 n < ck$.*

Proof By substitution into parts (a) and (b) of the Theorem. As in Corollary 1.1, Part (a) can be satisfied for any λ by a suitable K, and part (b) can be satisfied for any c, k for a suitable λ.

Corollary 1.3 *Theorem 1 with $r = R$ also holds for the Basic Mechanism realizing (k, k', R)-composable-associations.*

Proof The difference between standard and composable associations is that in the former the X_i are chosen randomly and independently, while in the latter they may be equated with a Y_j that was chosen randomly earlier. Since each Y_j can be equated with at most one X_i, the latter in the composable case can be viewed as independently randomly chosen, for all the purposes of the proof of the Theorem.

XIII. Realizing Associations by the Basic Mechanism: Limitations

The following Theorem gives some upper bounds on the capacity that can be achieved by the Basic Mechanism, for the case $k' \leq \frac{k}{8}$ and $k \geq 40$. We make two observations: First, part (a) together with the inequality $C < \eta_b^2 \nu$ from (b) imply that the capacity is indeed upper bounded by the default estimate $\nu = \frac{dn}{rk}$. Theorem 1 showed that this $\eta_b^2 \nu$ is achievable to constant factors. Second, the last bound $(\frac{d}{k})^2 (\frac{R}{r})$ implies that for the composability condition $R = r$, the capacity is bounded in terms of d and k, independent (!) of n and R.

Theorem 2 *Where $\nu = \frac{nd}{kr}$, $\eta_b = \frac{kn}{Rd}$, $k \geq 40$, and $\frac{Cr}{n} \geq 10$, for the Basic Mechanism to realize (k, k', R, r)-associations to capacity $C > 1$ on random graph $G_{n,d}$ it is necessary (a) for any k' that $\eta_b < 1$, and (b) for $k' \leq \frac{k}{8}$ that $C < \frac{n^3 k}{drR^2} = \eta_b^2 \nu \leq \frac{n^2}{Rr} \leq \nu \leq (\frac{d}{k})^2 (\frac{R}{r})$.*

Proof (A) Consider a single association $X_1 \to Y_1$. Consider the X_1 and Y_1 as already chosen, while the graph G is randomly generated. For each $x \in X_1$ and $y \in Y_1$ the probability that the edge (x, y) will be set high by the Basic Mechanism is $\frac{d}{n}$, namely the probability $\frac{d}{n}$ that (x, y) is present in G. Then the probability p that at least k edges incoming to y from the neurons X_1 are set high by the mechanism acting for $X_1 \to Y_1$ is the probability that in R independent Bernoulli trials (choices of $x \in X_i$) each with probability $\frac{d}{n}$ of success, that there are at least k successes, which is $p = \gamma^{\geq}(R, \frac{Rd}{n}, k) = \beta^{\geq}(R, \frac{Rd}{n}, \frac{kn}{Rd})$. We show that $\eta_b < 1$ by contradiction. Assume $\eta_b = \frac{kn}{Rd} \geq 1$. Then the expected number of successes in these Bernoulli trials will be $\frac{Rd}{n} \leq k$. Hence, by Proposition 2(ii) $p > \frac{3}{4}$ is not possible. However, condition (A) requires that this p exceed $1 - \frac{1}{2rC}$, which is at least $\frac{3}{4}$ if $C > 1$.

(B) Now consider that the C associations $X_i \to Y_i$, $1 \leq i \leq C$, have been realized with randomly chosen sets X_i of size R and Y_i of size r, and that the effectuating set E^* results. For condition (B) we need for every i that not many edges (x, y) with $x \in X_i$ and $y \notin Y_i$ be made effectuating by the $X_s \to Y_s$ with $s \neq i$. For contradiction we shall assume that $C = \frac{n^3 k}{drR^2}$.

For every neuron $y \in V$ define Q_y to be the set of indices s, with $1 \leq s \leq C$, such that $y \in Y_s$. It is easily seen, using Proposition 1(i), that, for any fixed y, $|Q_y| < \frac{Cr}{2n}$ (i.e. is less than a half of the expected size) with probability at most $p_1 = 1 - \beta^{\geq}(C, \frac{Cr}{n}, \frac{1}{2}) \leq e^{-\frac{Cr}{n}} (2e)^{\frac{Cr}{2n}} = (\frac{2}{e})^{\frac{Cr}{2n}}$. Hence, if $\frac{Cr}{2n} \geq 5$ then $p_1 \leq \frac{1}{4}$.

Consider i, Y_i, and $y \notin Y_i$ all as fixed, and G, X_i, and all the other X_s as randomly generated. Assume $|Q_y| \geq \frac{Cr}{2n}$. Choose a subset $Q' \subset Q_y$ of size exactly $\frac{Cr}{2n}$. Then for an $x \in V$ the probability that edge (x, y) is set high will be $p_{x,y} = \frac{R}{n} \frac{d}{n} (1 - (1 - \frac{R}{n})^{|Q_y|})$ where $\frac{R}{n}$ is the probability that $x \in X_i$, $\frac{d}{n}$ is the probability that the edge is in E, and $1 - (1 - \frac{R}{n})^{|Q_y|} \geq 1 - (1 - \frac{R}{n})^{\frac{Cr}{2n}}$ is the probability that $x \in X_s$ for at least one $s \in Q_y$. By Proposition 3(ii), if $\frac{RCr}{n^2} < 2$, then this $p_{x,y}$ is at least $\frac{R^2 dCr}{4n^4}$. But under the substitution $C = \frac{n^3 k}{drR^2}$, $\frac{RCr}{n^2} = \eta_b < 1$, (which is sufficient for $\frac{RCr}{n^2} < 2$) and the product $\frac{R^2 dCr}{4n^4}$ is $\frac{k}{4n}$. Now, as in the proof of Theorem 1, the multiplicative components in the expression for $p_{x,y}$ are negatively correlated, now we have them negatively lower correlated, and hence by Proposition 4 so are the $p_{x,y}$ for the various x for any one y. (For example, to see that the events $x \in X_i$ are negatively lower correlated we need to show that for any $S \subset V$, the probability $\text{Prob}(\bigwedge_{x \in S} x \notin X_i) \leq (\frac{n-R}{n})^{|S|}$. This follows inductively from the fact that $\text{Prob}(v_j \notin X_i | \bigwedge_{l<j} v_l \notin X_i) = \frac{n-R-j+1}{n-j+1} \leq \frac{n-R}{n}$ where $S = \{v_1, ..., v_{|S|}\}$.)

Hence the probability p_2 that there are at least k' successes is at least $\gamma^{\geq}(n, \frac{k}{4}, k')$ which by Proposition 1(i), is at least $1 - (\frac{2}{e})^{\frac{k}{8}}$ if $k' \leq \frac{k}{8}$. Hence $p_2 \geq \frac{3}{4}$ if $\frac{k}{8} \geq 5$.

Hence the probability that for the fixed i and fixed $y \notin Y_i$ there are at least $\frac{k}{8}$ effectuating edges from X_i to y is more than $(1-p_1) \times p_2 > \frac{1}{2}$. It follows that the expected number of $y \notin Y_i$ that have at least $\frac{k}{8}$ effectuating edges from X_i is greater than $\frac{1}{2}$, contradicting condition (B).

The last three inequalities follow by dividing $\frac{n^3 k}{drR^2}$ by η_b once, twice and thrice.

It follows that if $r = R$ then $C < \frac{d^2}{k^2}$. If also $d = O(n^{\frac{1}{2}})$ then the capacity is upper bounded linearly as $O(n)$ in terms of the number of nodes.

Corollary 2.1 *Under the conditions of Theorem 2 if $r = R$ then for $k' \leq \frac{k}{8}$, $C < (\frac{d}{k})^2$.*

Corollary 2.2 *Theorem 2 with $r = R$ also holds for the Basic Mechanism to realize $(k, \frac{k}{8}, R)$-composable associations.*

Proof Composable associations are a special case of associations.

XIV. Realizing Associations by the Expansive Mechanism: Possibilities

We now go on to consider whether the capacity limitations given above for the Basic Mechanism can be circumvented if one adopts the Expansive Mechanism. We will show that this is indeed the case, with the information theoretic limits being reachable now even in the composable setting, with $R = r$, which they were not for the Basic mechanism.

In the Expansive Mechanism one finds that the relay nodes impose some intricate probabilistic interdependences among the quantities to be analyzed. The upper and lower bound proofs are now more finely balanced, and for ease of analysis we shall be lax about the actual constants provided, which are for illustrative purposes and can be improved.

Corollary 3.1 will describe a case of great interest, with $d = D = n^{\frac{1}{2}}, r = R = 3k$, and $k = K \log_2 n$, for large enough constant K. We note that in this parameter regime the conditions (i) - (viii) of Theorem 3 are constraints that can be satisfied for large enough values of n.

Theorem 3 *For all positive integers $C > 1, n, d, D, k, R, r$ and k' where also $c = \frac{k'}{k}, \nu = \frac{nd}{kr}$ and $\eta_e = \frac{kn}{RdD}$, if (a) $C < \frac{1}{2r} e^{k(\frac{1}{2\eta_e} - 1)} (2\eta_e)^k$, and (b) $C \leq \frac{ckn^3}{128rR^2dD^2} = \frac{c}{128}\eta_e^2 \nu$, then on random graph $G_{n,d,D}$ the Expansive Mechanism realizes (k, k', R, r)-associations to capacity C, provided that (i) $DR \leq n$, (ii) $\eta_e \leq \frac{1}{2}$, (iii) $(\frac{e}{4})^D \leq \frac{1}{10n}$, (iv) $(\frac{e}{4})^{\frac{Cr}{n}} \leq \frac{1}{10n}$, (v) $(\frac{e}{4})^{\frac{ckn}{8d}} \leq \frac{1}{10nC}$, (vi) $(\frac{e}{D})^{\frac{2CrR^2D}{n^2}} \leq \frac{1}{10nC}$, (vii) $(\frac{e}{4})^{\frac{ck}{2}} \leq \frac{1}{10nC}$ and (viii) $8CrR^2 \leq n^2 D$.*

Proof (A) Consider an arbitrary i with $1 \leq i \leq C$, and the choice of X_i, Y_i as fixed, and G as randomly generated from $G_{n,d,D}$. Consider a $y \in Y_i$. Let M_y be the set of relay nodes that have connections to y. Let I_y be the set of basis nodes with edges to members of M_y. Then the probability p_u that an edge (u, y), where u is a fixed relay node, gets a high setting by the association $X_i \to Y_i$ is the product of the probability that (u, y) is an edge of G, and the probability that at least one of the nodes $x \in X_i$ is connected to u. In other words $p_u = \frac{d}{n}(1 - (1 - \frac{D}{n})^R)$. This by Proposition 3(ii) is no less than $\frac{RDd}{2n^2}$ provided $\frac{RD}{n} \leq 1$, a condition guaranteed by (i). Hence, the probability that at least k edges incoming to y are set high by the association $X_i \to Y_i$ is lower bounded by the probability that in n independent Bernoulli trials (corresponding to the choices of u) each with probability at least $\frac{RDd}{2n^2}$ of success, that there are at least k successes, which is at least $\gamma^{\geq}(n, \frac{RDd}{2n}, k)$ = $\beta^{\geq}(n, \frac{RDd}{2n}, \frac{2kn}{RdD})$. If $\eta_e \leq \frac{1}{2}$, as guaranteed by (ii), then $\alpha = \frac{2kn}{RdD} \leq 1$ and Proposition 1(i) applies, giving $\beta^{\geq}(n, \frac{RDd}{2n}, \frac{2kn}{RdD}) \geq 1 - e^{\frac{-RdD}{2n} + k} (\frac{2kn}{RdD})^{-k}$. Hence for $\beta^{\geq}(n, \frac{RDd}{2n}, \frac{2kn}{RdD})$ to exceed $1 - \frac{1}{2rC}$, and so realize condition (A), it is sufficient that $2rC < e^{\frac{RdD}{2n}} (\frac{eRdD}{2kn})^{-k}$, or that $C < \frac{1}{2r} e^{\frac{RdD}{2n} - k} (\frac{2kn}{RdD})^k$.

(B) Let F be the event that the following both hold: every vertex in the basis layer has outdegree at most $2D$, and for every basis layer vertex y, $|\{s | y \in Y_s\}| \leq \frac{2Cr}{n}$. It is easily seen, exactly as in Theorem 1, that the two parts have probabilities lower bounded respectively by (a) $1 - p_1$ where $p_1 = \beta^{>}(n, D, 2D) < (\frac{e}{4})^D$, and by (b) $1 - p_2$ where $p_2 = \beta^{>}(C, \frac{Cr}{n}, \frac{2Cr}{n}) < (\frac{e}{4})^{\frac{Cr}{n}}$. Hence if conditions (iii) and (iv) hold then the union bound can be applied to the p_1 and p_2 derived, for each of the n vertices, to obtain that the probability that F fails to hold is less than $\frac{1}{5}$.

We need that all the C associations $X_i \to Y_i$ $(1 \leq i \leq C)$ be realized such that the firing of any one X_i should not cause many spurious firings of ys such that $y \notin Y_i$. Such spurious firings are caused when there are many relay nodes u each connected from a node in X_i and also connected to y, such that the edge (u, y) is set high by some association $X_s \to Y_s$ with $s \neq i$. A necessary and sufficient condition for the edge (u, y) to be set high by $X_s \to Y_s$ for some $s \neq i$ is that "$y \in Y_s$, u is connected from some node v in X_s, and $(u, y) \in E$". There will be two cases to distinguish, according to whether $v \in X_s \cap X_i$ or $v \in X_s - X_i$.

Suppose that F holds, and consider a fixed i, Y_i, and $y \notin Y_i$. Let U_i be the set of relay nodes to which there are connections from X_i, $W_{i,y}$ be the set to which there is an edge from some basis layer node $v \in X_s - X_i$ for some s with $y \in Y_s$, and $T_{i,y} \subset U_i$ be the set to which there is an edge from some basis layer node $v \in X_i \cap X_s$ for some s with $y \in Y_s$. We shall upper bound the size of the set of relay nodes $R_{i,y} = (W_{i,y} \cup T_{i,y}) \cap U_i = Z_{i,y} \cup T_{i,y}$ where $Z_{i,y} = (U_i \cap W_{i,y})$.

We first upper bound $|Z_{i,j}|$, corresponding to $v \in X_s - X_i$. Now $|U_i| \leq 2RD$ since there are R vertices in X_i and, if F holds, each is connected to at most $2D$ relay nodes. Also, if F holds, $|W_{i,y}| \leq \frac{4DRCr}{n}$, since then there are at most $\frac{2Cr}{n}$ choices of s for which $y \in Y_s$, and for each such s there are R members of X_s and each is connected to at most $2D$ relay nodes.

We observe that the choice of $W_{i,y}$ is independent of the choice of U_i since the former depends on xs outside of X_i, on the edges of G incident to those xs, and on the membership of those xs in those other X_ss. We need to upper bound the size $|Z_{i,y}|$ of their intersection $(U_i \cap W_{i,y})$.

Applying Proposition 5(i) with $A = U_i$ and $B = W_{i,y}$ gives the probability that $|Z_{i,y}| \geq \frac{32D^2R^2Cr}{n^2}$ is at most $(\frac{e}{4})^{\frac{16D^2R^2Cr}{n^2}}$. Suppose that $C = \frac{ckn^3}{128rR^2dD^2}$. Then it follows that this latter probability $(\frac{e}{4})^{\frac{16D^2R^2Cr}{n^2}}$ is at most $p_3 = (\frac{e}{4})^{\frac{ckn}{8d}}$ and that except with that probability, $|Z_{i,y}| < \frac{ckn}{4d}$.

We now upper bound $|T_{i,y}|$, corresponding to the case of $v \in X_i \cap X_s$ and show that $|T_{i,y}| \leq \frac{32D^2R^2Cr}{n^2}$ with a similar overwhelming probability: If F holds the probability that for a fixed $v \in X_i$ it is the case that $v \in X_s$ for at least one of the at most $\frac{2Cr}{n}$ values of s such that $y \in Y_s$ is at most $1 - (1 - \frac{R}{n})^{\frac{2Cr}{n}}$ which by Proposition 3(i) is at most $\frac{2CrR}{n^2}$, provided $\frac{2Cr}{n} \geq 1$, which is implied by (iv). For different vs these probabilities are not independent, but they are negatively upper correlated, as in the proof of Theorem 1. Hence, by Proposition 1(v) and Proposition 1(ii), the probability that there are more than $\frac{2R^2CrD}{n^2}$ such vs in X_i is upper bounded by $\beta^{\geq}(R, \frac{2CrR^2}{n^2}, D)$, which by Proposition 1(ii) is at most $(\frac{e}{D})^{\frac{2CrR^2D}{n^2}}$ provided $D \geq 1$, which is guaranteed by assumption. But this $p_4 = (\frac{e}{D})^{\frac{2CrR^2D}{n^2}} \leq \frac{1}{10nC}$ if condition (vi) holds. Hence, since each such $v \in X_i$ has degree at most $2D$ if F holds, it follows that $|T_{i,y}| \leq \frac{4D^2R^2Cr}{n^2}$ which is less than $\frac{ckn}{32d}$ if $C \leq \frac{ckn^3}{128R^2rdD^2}$.

We now combine the analyses of the two cases of $v \notin X_i$ and $v \in X_i$ to get $|R_{i,y}| \leq |Z_{i,y}| + |T_{i,y}| \leq \frac{9ckn}{32d} \leq \frac{ckn}{2d}$, except with probability $p_3 + p_4$. Then the number $|t_{i,y}|$ of relay nodes from which edges to y will get high values by the associations $s \neq i$ will be the fraction of $R_{i,y}$ from which edges to y exist in G, the probability of any one such edge being $\frac{d}{n}$. Hence, $|t_{i,y}|$ will exceed twice the expectation, namely, ck, with probability at most $\gamma^{>}(\frac{ckn}{2d}, \frac{ck}{2}, ck)$ which by Proposition 1(iv) is less than $p_5 = (\frac{e}{4})^{\frac{ck}{2}}$, which is bounded by (vii).

Now if we make $p_1 \leq \frac{1}{10n}, p_2 \leq \frac{1}{10n}, p_3 \leq \frac{1}{10nC}, p_4 \leq \frac{1}{10nC}$ and $p_5 \leq \frac{1}{10nC}$, then by the union bound taken over y for p_1 and p_2, and over y and i for p_3, p_4 and p_5, the probability of any error occurring by any of these five means is less than $\frac{1}{2}$. This establishes property (B) of Definition 1.

Corollary 3.1 *For any constant $c > 0$, there exist λ, K such that for $d = D = n^{\frac{1}{2}}, r = R = 3k, k = K \log_2 n$, capacity $C = \frac{\lambda n^{\frac{3}{2}}}{,} (\log_2 n)^2$ can be achieved.*
Proof By substitution into parts (a) and (b) of the Theorem. Part (a) can be satisfied for any λ by a suitable K. Part (b) can be satisfied for any c, K by a suitable λ.

Corollary 3.2 *Theorem 3 with $r = R$ also holds for the Expansive Mechanism realizing (k, k', R)-composable-associations.*
Proof The argument of Corollary 1.3 holds here also.

XV. Realizing Associations by the Expansive Mechanism: Limitations

The simple information theoretic $C \leq \tilde{O}(nd)$ upper bound from Section III applies to the Expansive Mechanism just as it does for the Basic Mechanism, since in both the only modifiable synapses are at the n basis nodes of expected indegree d. Hence, for the composable case of interest described in Corollary 3.1, namely $d = D = n^{\frac{1}{2}}$, the Expansive Mechanism achieves optimal capacity $\tilde{\Theta}(n^{\frac{3}{2}})$.

Here we shall give a lower bound that is stronger than the information theoretic bound in that it applies to a broader range of values of d, D. Part (a) shows that for the expansive efficacy $\eta_e = \frac{kn}{RdD}$, $\eta_e < 1$ again must hold without condition. Second, a lower bound is given that applies to wide ranges of the parameters, subject to some technical constraints (i) - (viii). Part (b) shows that the default estimate $\nu = \frac{dn}{rk}$ is still an upper bound on the capacity up to constant factors. Note that Corollary 4.1 will show that for $r = R$, there is an upper bound on capacity independent of n, R and r, but this now depends on the degrees as d^2D, rather than as the d^2 upper bound of Corollary 2.1, or the nd information theoretic bound.

Theorem 4 *For all positive integers $n > 1, d, D, k, R$ and r, where $\nu = \frac{nd}{kr}$ and $\eta_e = \frac{kn}{RdD}$, for the Expansive Mechanism to realize (k, k', R, r)-associations to capacity $C \geq 2$ on random graph $G_{n,d,D}$ it is necessary (a) for any k' that $\eta_e < 1$, and (b) for $k' \leq \frac{k}{16}$, that $C < \frac{128n^3k}{D^2drR^2} = 128\eta_e^2\nu \leq 128\frac{n^2}{DRr} \leq 128\nu \leq 128(\frac{d}{k})^2\frac{RD}{r}$, provided that (i) $(\frac{2}{e})^{\frac{D}{2}} \leq \frac{1}{12n}$, (ii) $(\frac{2}{e})^{\frac{Cr}{2n}} \leq \frac{1}{12n}$, (iii) $DR \leq n$, (iv) $(\frac{2}{e})^{\frac{DR}{4}} \leq \frac{1}{12C}$, (v) $CrDR \leq 2n^2$, (vi) $(\frac{2}{e})^{\frac{CrR}{64n}} \leq \frac{1}{12nC}$, (vii) $(\frac{2}{e})^{\frac{kn}{8d}} \leq \frac{1}{12nC}$, (viii) $(\frac{2}{e})^{\frac{k}{16}} \leq \frac{1}{12nC}$.*

Proof (A) Consider the first component $X_1 \rightarrow Y_1$ of the associations, and consider X_1, Y_1 as fixed. Consider a $y \in Y_1$. Now consider G as randomly chosen. Then the probability p that an edge (u, y), where u is a fixed relay node, gets a high setting is the product of the probability that (u, y) is an edge of G, and the probability that at least one of the basis nodes in X_1 is connected to u. In other words $p = \frac{d}{n}(1 - (1 - \frac{D}{n})^R)$. This, by Proposition 3(i) is no more than $\frac{RDd}{n^2}$. Hence, the probability that at least k edges incoming to y are set high by firing X_1 is the probability, in n independent Bernoulli trials (corresponding to the choices of u) each with probability no more than $\frac{RDd}{n^2}$ of success, that there are at least k successes, which is no more than $\gamma^{\geq}(n, \frac{RDd}{n}, k) = \beta^{\geq}(n, \frac{RDd}{n}, \frac{kn}{RdD})$. Hence for this to exceed $1 - \frac{1}{2rC} \geq \frac{3}{4}$ and so achieve condition (A) we need, by Proposition 2(ii), that $\frac{kn}{RDd} = \eta_e < 1$.

(B) Suppose all the C associations $X_s \to Y_s$ $(1 \leq s \leq C)$ can be realized such that the firing of X_i does not cause many spurious firings of ys such that $y \notin Y_i$. Such spurious firings are caused when a relay node u connected from a node in X_i is also connected to a $y \notin Y_i$, such that the edge (u, y) is set high by some association $X_s \to Y_s$. A necessary and sufficient condition for the edge (u, y) to be set high by $X_s \to Y_s$ is that $y \in Y_s$, u is connected from some node in X_s, and $(u, y) \in E$.

Let F be the event that the following both hold: (a) every vertex in the basis layer has outdegree at least $\frac{D}{2}$, and (b) for every basis layer vertex y and every s with $1 \leq s \leq C$, if Q_y is defined as $\{s | y \in Y_s\}$ then $|Q_y| \geq \frac{Cr}{2n}$. Now, it is easily seen using Proposition 1(i) that the two parts have probabilities lower bounded by (a) $1 - p_1$ where $p_1 = 1 - \beta^{\geq}(n, D, \frac{D}{2}) \leq (\frac{2}{e})^{\frac{D}{2}}$, and (b) $1 - p_2$ where $p_2 = 1 - \beta^{\geq}(C, \frac{Cr}{n}, \frac{Cr}{2n}) \leq (\frac{2}{e})^{\frac{Cr}{2n}}$. Hence if conditions (i) and (ii) hold then applying these to p_1, p_2 for all the n values of y gives that F holds except with probability $\frac{1}{6}$.

Now suppose that F holds, and consider a fixed i, Y_i, and $y \notin Y_i$. We shall lower bound the size of the set of relay nodes $Z_{i,y} = U_i \cap W_{i,y}$, where U_i is the set of relay nodes to which there are connections from X_i, and $W_{i,y}$ is the set of relay nodes to which there is an edge from some basis layer node v such that $v \in X_s$ for some s with $y \in Y_s$.

We first lower bound $|U_i|$. The probability that relay node u is not in U_i is the probability that every edge from X_i to u is missing, which is $(1 - \frac{D}{n})^R$. We apply Proposition 3(ii) to get $(1 - \frac{D}{n})^R < 1 - \frac{DR}{2n}$, provided $\frac{DR}{n} \leq 1$, which is guaranteed by assumption (iii). Hence u is present in U_i with probability at least $\frac{DR}{2n}$, independently for each u. Hence $|U_i|$ is less than half the mean, $\frac{DR}{4}$ with probability at most $p_3 = 1 - \gamma^{\geq}(n, \frac{DR}{2}, \frac{DR}{4}) \leq (\frac{2}{e})^{\frac{DR}{4}}$, by Proposition 1.3.

Now we lower bound $|W_{i,y}|$. Let $H = \{v | v \in X_s \text{ for some } s \in Q_y, \text{but } v \in X_i\}$. Since, $|Q_y| \geq \frac{Cr}{2n}$, the probability that a vertex $v \notin X_i$ belongs to H is at least $1 - (1 - \frac{R}{n})^{\frac{Cr}{2n}}$ which by Proposition 3(ii) is at least $\frac{CrR}{4n^2}$, provided $\frac{CrR}{2n^2} \leq 1$, which follows from (v). From (ii), $\frac{Cr}{2n} \geq 2$. But if also $\frac{Cr}{2n} \leq \frac{n}{R}$ (from (v)) then $R \leq \frac{n}{2}$. Hence the probability that $|H| \leq \frac{CrR}{16n}$ is no more than $1 - \gamma^{\geq}(\frac{n}{2}, \frac{CrR}{8n}, \frac{CrR}{16n}) \leq (\frac{2}{e})^{\frac{CrR}{16n}}$ by Proposition 1.3, and, in turn, by (vi), is then upper bounded by $\frac{1}{12nC}$.

Now each member of $W_{i,y}$ arises as one of at least $\frac{D}{2}$ random choices for each of the at least $\frac{CrR}{16n}$ choices H. Hence the probability p_4 that $|W_{i,y}| < \frac{CrRD}{64n}$ is at most $1 - \gamma^{\geq}(n, \frac{CrRD}{32n}, \frac{CrRD}{64n}) \leq (\frac{2}{e})^{\frac{CrRD}{64n}}$, which by (vi) is at most $\frac{1}{12nC}$.

It remains to lower bound the size of the intersection $|Z_{i,y}| = |W_{i,y} \cap U_i|$, where $|W_{i,y}| \geq \frac{CrRD}{64n}$ and $|U_i| \geq \frac{DR}{4}$. We apply Proposition 5(ii) with $A = U_i$ and $B = W_{i,y}$ to get that the probability that $|Z_{i,y}| \leq \frac{D^2R^2Cr}{1024n^2}$ is less than $(\frac{2}{e})^{\frac{D^2R^2Cr}{1024n^2}}$. Then if $C \leq \frac{128kn^3}{rR^2dD^2}$ it follows that this latter probability $(\frac{2}{e})^{\frac{D^2R^2Cr}{1024n^2}}$ is at most $p_5 = (\frac{2}{e})^{\frac{kn}{8d}}$ (which by (vii) is at most $\frac{1}{12nC}$) and that except with that probability $|Z_{i,y}| \geq \frac{kn}{8d}$.

But if $|Z_{i,y}| \geq \frac{kn}{8d}$, then the number $|t_{i,y}|$ of relay nodes from which edges to y will get high values by the associations $s \neq i$ will be the number of the $Z_{i,y}$ from which edges to y exist in G, the probability of any one such edge being $\frac{d}{n}$. Hence, $|t_{i,y}|$ will exceed half the expectation, namely, $\frac{k}{16}$, with probability at least $\beta^{\geq}(\frac{kn}{8d}, \frac{k}{8}, \frac{k}{16}) \leq 1 - (\frac{2}{e})^{\frac{k}{16}}$, by Proposition 1(i). Hence, the ys will be $\frac{k}{16}$-excited except with probability at most $p_6 = (\frac{2}{e})^{\frac{k}{16}}$.

Now if conditions (i), (ii), (iv), (vi), (vii) and (viii) make $p_1, p_2 \leq \frac{1}{12n}$, $p_3 \leq \frac{1}{12C}$, and $p_4, p_5, p_6 \leq \frac{1}{12nC}$, so that by the union bound the probability of any error through any occurrence of any of these events is at most $\frac{1}{2}$. This establishes property (B) of Definition 1.

The last three inequalities follow by dividing $\frac{128n^3k}{D^2drR^2}$ by η_e repeatedly and using (a).

Corollary 4.1 *Under the conditions (i) - (viii) of Theorem 4, if $r = R$ then for $k' < \frac{k}{16}$ the capacity is limited by $C \leq \frac{128d^2D}{k^2}$.*
Proof By substituting $r = R$ in the statement of the Theorem.

Corollary 4.2 *Theorem 4 with $r = R$ also holds for the Expansive Mechanism when realizing $(k, \frac{k}{16}, r)$-composable-associations.*
Proof Composable associations are a special case of associations.

XVI. Conclusion

Generic concepts of distributed representations, sparse representations, grandmother cells have all been discussed extensively in the neuroscience literature. Our approach may be viewed as providing concrete semantics and quantitative analysis for these otherwise imprecise notions.

Our assumption that every concept is represented by the same exact number of neurons was already made in the Willshaw net literature [WBLH69]. However, in the current context, where we seek a cognitively adequate set of primitives that includes hierarchical memorization, this assumption needs some justification, because the simplest implementations of the latter do not maintain such a fixed number of neurons allocated to each concept [V94, G03]. However, this assumption is made reasonable by the more recent result that simple feedforward networks can indeed stabilize the numbers allocated [V12]. Without that stability assumption the analysis becomes more difficult. Note that the simulation results in [FV09] are incomparable with the results here in that a broader set of primitives is implemented there, and no stability enforced.

As a purely speculative note we observe that for the Expansive Mechanism our asymptotic positive result is quite delicate, and its advantages over the Basic Mechanism may occur only for large cortices, such as those of humans. This may be correlated with the question of whether our functionality combination (1) - (6) described in the Introduction is something most particular to humans. While these considerations all motivate the current study, the possibility that the Basic Mechanism is sufficient to explain human performance has not been ruled out. If, as we expect, associations need to be composed to only very limited depth, requirements laxer than our Definitions 1 and 2 may suffice.

An important empirical question is whether evolution has discovered the Expansive Mechanism. There are some experimental indications of the existence of strong synaptic connections, including in human cortex [MOK+08, DB15, MRB+16]. However, the question of whether any are used in a manner similar to our mechanisms remains unresolved.

Acknowledgment

I am grateful to Daniel Moroz for his comments on an earlier version of this paper. This research was supported in part by National Science Foundation grant CCF-1509178 and DARPA grant W911NF-16-1-0561, and conducted in part at the Microsoft New England Research and Development Center, Cambridge, MA.

References

[BK08] Beal, J. & Knight, T.F. (2008). Analyzing composability in a sparse encoding model of memorization and association. *Proc. 7th IEEE International Conference on Development and Learning (ICDL 2008)*,180–185.

[BK98] Bosch, H. & Kurfess, F. J. (1998). Information storage capacity of incompletely connected associative. *Neural Networks*, *11*:5, 869–876.

[BW92] Buckingham, J. and Willshaw, D. (1992). Performance characteristics of the associative net. *Network*, *3*: 407–414.

[DA03] Dayan, Peter & Abbott, L.F. (2003). *Theoretical Neuroscience: Computational and Mathematical. Modeling of Neural Systems*. MIT Press,Cambridge, MA.

[DB15] Doron, Guy & Brecht, Michael S. M. (2015). What single-cell stimulation has told us about neural coding *Phil Trans Roy Soc B:*, *370* 20140204.

[ELS14] Einarsson, H., Lengler, J. and Steger, A. (2014). A high-capacity model for one shot association learning in the brain *Frontiers in Computational Neuroscience*,*8*,140.

[FV09] Feldman, V. & Valiant, L. G. (2009). Experience induced neural circuits that achieve high capacity. *Neural Computation*, *21*, 2715–2754.

[G03] Gerbessiotis, A.V. (2003). Random graphs in a neural computation model. International Journal of Computer Mathematics, *80*, 689–707.

[GW97] Graham, B. & Willshaw, D. (1997). Capacity and information efficiency of the associative net. *Network: Comput. Neural Syst.*, *8*, 35–54.

[H49] Hebb, D.O. (1949). *The Organization of Behavior*. Wiley, New York, NY.

[H82] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *PNAS*, *79*(8)2554–2558.

[KPS03] Knoblauch, A., Palm, G., & Sommer, F.T. (2010). Memory capacities for synaptic and structural plasticity. *Neural Computation*, *94*(1)33–45.

[MOK+08] Molnar G, Olah S, Komlosi G, Fule M, Szabadics J, Varga C, Barzo P, Tamas G. (2008) Complex events initiated by individual spikes in the human cerebral cortex. *PLoS Biol.* *6*, e222. (doi:10.1371/journal.pbio. 0060222)

[MRB+16] Molnar G, Rozsa, M. S, Baka, J., Holderith, N., Barzo, P., Nusserr, Z. Tamas G. (2016) Human pyramidal to interneuron synapses are mediated by multi-vescular release and multiple docked vesicles. *eLIFE 5*, e18167. (doi:10.7554/eLife.18167)

[PS97] Panconesi, A. & Srinivasan, A. (1997) Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds. *SIAM J. COMPUT.*, *26*:2, 350–368.

[PV15] Papadimitriou, C.H. and Vempala, S. S. (2015). Cortical Learning via Prediction. *J. Mach. Learn. Res*, *40*, 1-21.

[S96] Sanders, P. (1996). On the competitive analysis of randomized static load balancing. *RANDOM*, *40*, 1-21.

[V94] Valiant, L.G. (1994). *Circuits of the Mind*. New York, NY.

[V05] Valiant, L.G. (2005). Memorization and association on a realistic neural model. *Neural Computation*, *17*, 527–555.

[V06] Valiant, L.G. (2006). A quantitative theory of neural computation. *Biological Cybernetics*, *95*:3, 205–211.

[V12] Valiant, L.G. (2012). The hippocampus as a stable memory allocator for cortex. *Neural Computation*, *24*, 2873–2899.

[WBLH69] Willshaw, D. J., Buneman, O. P. and Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature*, *222*, 960–962.

[W71] Willshaw, D. (1971). Models of distributed associative memory. PhD Thesis, University of Edinburgh.