

available at www.sciencedirect.comwww.elsevier.com/locate/brainres**BRAIN
RESEARCH**

Research Report

The hypergeometric connectivity hypothesis: Divergent performance of brain circuits with different synaptic connectivity distributions

Andrew C. Felch, Richard H. Granger*

Psychological and Brain Sciences Department, and Neukom Institute for Computational Science, Dartmouth College, 6207 Moore Hall, Hanover, NH 03755, USA

ARTICLE INFO

Article history:

Accepted 9 June 2007

Keywords:

Brain circuit model
Synaptic distribution
Thalamocortical model
Development
Computational neuroscience

ABSTRACT

The development of connectivity among brain networks (e.g., thalamocortical, cortico-thalamic, cortico-cortical) proceeds via a combination of axon and dendrite growth followed by a later process of synaptic pruning [Purves, D., Lichtman, J.W., 1980. Elimination of synapses in the developing nervous system. *Science*, 210, 153–157; Oppenheim, R.W., 1991. Cell death during development of the nervous system. *Annual Review of Neuroscience*, 14 (1), 453–501.; Oppenheim, R., Qin-Wei Y., Prevet D., Yan Q., 1992. Brain-derived neurotrophic factor rescues developing avian motoneurons from cell death. *Nature*, 360, 755–757]. Sparse synaptic distribution (i.e., the low probability (<0.1) of contact among neurons; [Braitenberg, V., Schüz, A., 1998. *Cortex: Statistics and geometry of neuronal connectivity*: Springer Berlin.] can conform to any of a range of connectivity patterns with different distributional characteristics; and different distribution patterns can yield networks with very different functional properties. We rigorously investigate a range of different connectivity characteristics, and show that different synaptic distributions can substantially affect the functional capabilities of the resulting networks. In particular, we provide formal measures of information loss in transmission from one set of neurons to another as a function of synaptic distribution, and show a set of empirical cases with different information-theoretic utility. We characterize the trade-offs among utility and costs, and their dependence on different classes of developmental strategies by which axons from one cell group are “assigned” to synapses on dendrites from a target cell group. It is shown that hypergeometric distributions minimize a range of measured costs, compared to competing synaptic distributions. It is also found that the divergent performance among differently organized brain circuits expands with brain size, rendering the effects increasingly consequential for big brains. In summary, we propose that the characteristics of hypergeometric connectivity provide a coherent explanatory hypothesis of a range of developmental and anatomical data.

© 2007 Elsevier B.V. All rights reserved.

* Corresponding author.

E-mail address: Richard.Granger@Dartmouth.edu (R.H. Granger).

0006-8993/\$ – see front matter © 2007 Elsevier B.V. All rights reserved.

doi:10.1016/j.brainres.2007.06.044

1. Introduction

A prominent feature of brain networks is their sparse connectivity; the probability of one projection neuron making synaptic contact with a distant target is less than 0.01 (Braitenberg and Schüz, 1998) and even for close neighbors, the probabilities rise only to the range of about 0.1. This sparseness presumably arises in part to conform with scaling constraints: as the number of neurons increases, the number of synapses required to connect every neuron to every other would rise prohibitively as the square of the number of neurons; sparseness effectively prevents the system from growing without bound. Yet the mathematics of sparse matrices differ notably from those of dense matrices, in some instances enabling computations that depend on sparseness; and it is possible that some of the still-elusive computational power of the brain arises in part from its connectivity characteristics (see, e.g., Kilborn et al., 1996).

If not all neurons are connected, which become connected to which? Do some axons make more synapses than others; do some dendrites receive more synapses than others? Developmental neuroscience has identified many of the fundamental mechanisms by which neural connections grow, providing constraints on possible connection patterns; additional constraints may be found via simulation and analysis. In the modeling exercises presented here, we demonstrate that these deceptively simple questions have an unexpectedly large effect on the computational powers of the resulting networks. That is, circuits wired according to just slightly different specifications exhibit very different computational characteristics, when embedded in circuit designs characteristic of mammalian telencephalon. Moreover, and equally surprising, the differences among these circuits diverge strongly as the size of the circuits increases—thus the choice of wiring pattern differentially affects the circuits in big brains far more than in small brains, i.e., there are larger advantages in larger brains. The findings converge on a particular pattern of connectivity which we forward as a hypothesis of the synaptic connectivity distribution that occurs in mammalian brain circuits.

The vast majority of neural network research has thus far focused almost entirely on fully connected models. Even within the realm of partial connectivity modeling, not much importance has been placed on random static synapse distribution patterns (Elizondao et al., 1995), though this is currently a best approximation of human (and all mammalian) brain networks. Studies of ontogenic methods that allow creation and destruction of synapses (Fiesler, 1994) over the scale of hours are unlikely to account for the rapidly induced (seconds) learning that typically occurs in “everyday” adult human fact learning. Substantial evidence suggests that, after synaptic growth during development, a “static” distribution of synapses is arrived at in the adult, and that even if further growth and pruning occur, significant learning can proceed just via local modifications (long-term potentiation) to existing synapses (see, e.g., Morris et al., 1986; Bannerman et al., 1995; Saucier and Cain, 1995; Rioult-Pedotti et al., 2000).

It might have been the case that many different synaptic distribution variations would suffice to enable adult learning, and this belief is often stated in the literature, e.g.:

“The form of sparse connectivity chosen in the present article is such that each neuron receives exactly the same number of inputs. This was done here for the sake of simplicity.” (Brunel, 2000)

In contrast, we show here that different synaptic distributions render both quantitative and qualitative performance differences in the resulting networks.

We focused on the following four logical possible distribution patterns of synapses: (i) that synapses are more tightly distributed (i.e., with lower variance) across axons than across dendrites; (ii) that synapses are more tightly distributed across dendrites than across axons; (iii) that synapses are not tightly distributed across either axons or dendrites; or (iv) that synapses are tightly distributed across axons and across dendrites.

In the following empirical tests, we idealize these four distribution patterns such that tight distributions approximate uniform distributions whereas looser distributions approximate arbitrary distributions.

Correspondingly, four different classes of connectivity models were studied, corresponding to idealizations of these

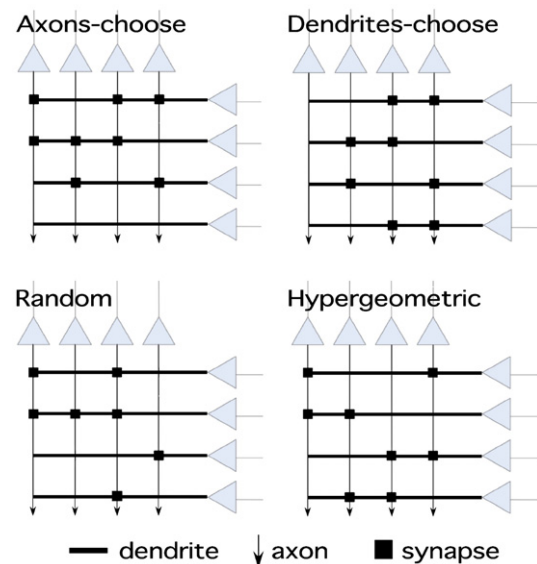


Fig. 1 – Four logical patterns of distribution of synapses among axons and dendrites. Axons-choose: Synapses are distributed uniformly across axons (vertical arrows), but arbitrarily across the target dendritic field (horizontal lines). Dendrites-choose: Synapses are distributed uniformly across target dendrites but arbitrarily across incoming axons. Random: Synapses are arbitrarily distributed across incoming axons and target dendrites. Hypergeometric: Each incoming axon produces some fixed number of synapses, and each target dendrite receives a fixed number (not necessarily the same number) of synapses from those axons, rendering a uniform distribution of synapses across axons and across dendrites (see text).

four logically possible ways that axon and dendrite populations can form synaptic distributions (see Fig. 1):

Axons-Choose: Each axon incoming to a target dendritic field produces exactly S synapses, distributed arbitrarily among target neurons.

Dendrites-Choose: Each target cell receives input from exactly S input axons.

Random: Synapses between source and target neurons are assigned probabilistically, and allowing possible redundancy.

Hypergeometric: Axons of each source neuron produce a fixed number S_1 of synapses, and each target neuron receives a fixed number S_2 of synapses.

Probabilistic or random allotment is most widely used in models; Dendrites-Choose and Axons-Choose are logical variants. Hypergeometric connectivity may at first seem difficult to achieve, but may approximate the actual, surprisingly regular, distributions of synapses in real brain circuits (see, e.g., Braitenberg and Schüz, 1998); close approximations to hypergeometric distributions may readily be achieved via initial overgeneration of synaptic contacts followed by selective die-off, concordant with what is actually observed during brain development (Purves and Lichtman, 1980; Oppenheim, 1991; Oppenheim et al., 1992; Schütze, 1993; Bose, 2003).

Each of these four synaptic distribution patterns was implemented, and four different metrics were employed to measure the functional characteristics of each class of network. The canonical network design used for these tests was a “ k -winner-take-all” (kWTA) network, often cited as a possible approximation of naturally occurring networks, such as typical neocortical local circuits, containing excitatory feedforward and inhibitory feedback elements (Grossberg, 1976; Amari and Arbib, 1977; Yuille and Grzywacz, 1989; Coultrip et al., 1992; Kilborn et al., 1996; Thomson and Bannister, 2003).

In the results that follow, five experiments using four metrics were chosen to investigate these network patterns. Due to the breadth of neural network models that exist in the literature, the results will apply differentially to different models with different architectures. While the results will certainly not be general enough to substitute for running experiments directly on a particular model, they indicate general tendencies that may be useful in the absence of specific connectivity-centric experiments. Indeed one goal of these experiments is to determine whether such study is likely to be important.

2. Results

2.1. Experiment 1: information maintenance

The first test measures information maintenance (IM) and was derived from the work of Adelsberger-Mangan and Levy (1992). IM was measured by presenting a network with 2000 different random inputs, and the number of unique outputs was counted. Collisions occurred when two different inputs caused the same output. A low number of collisions indicate that the

network was capable of relaying information efficiently. Such networks are considered crucial for transmitting information from one cortical area to another, either directly through cortico-cortical connections or indirectly through a convergent-to-divergent nonspecific thalamus relay. IM is related to the theory of “information maximization” (Bell and Sejnowski, 1995), but information maximization applies only when there is some correlation between the inputs. Information maintenance experiments may apply more broadly than those of information maximization because no assumptions are made as to the types of correlations that exist between inputs. Fig. 2 shows the results of this analysis (see Section 4.1 for details).

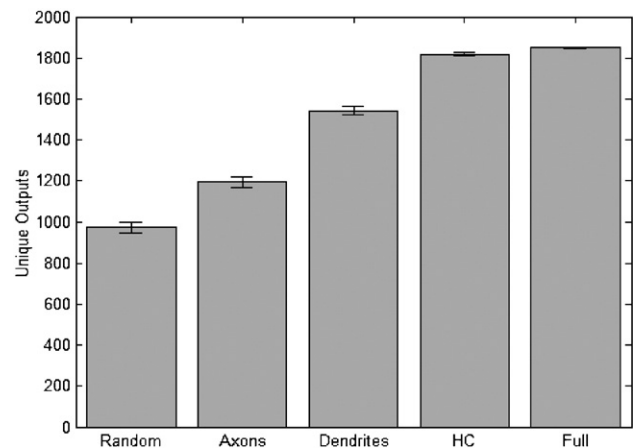


Fig. 2 – Performance of five classes of network in an information-maintenance (IM) task. Shown are the number of unique outputs (y axis; means and standard errors) for each of four sparse network classes plus a fully connected network. To measure IM, collisions were made likely by using 20 output neurons with 10 active (see text). Connection density was 10% with 20 input neurons and 50% activity density in both input and output layers. Each test was run 100 times and the results were averaged. Initial studies showed that allowing duplicate connections usually led to inferior performance when compared with a non-redundant counterpart model. Thus, only the random synapse distribution pattern was allowed to form duplicate connections. It can be seen that, in this task, Random and Axons-Choose patterns display the poorest performance, and much information is lost in these networks. The Dendrites-Choose pattern (Gotelli and Entsminger, 2003) also loses some information. Hypergeometric connectivity (HC) enables sparsely connected networks to perform as well as networks with full connectivity. It should be noted that similar results were achieved using 20% input and 20% output network activity densities. In order of the x-axis, the results were [440 600 590 900 990] (rounded to two significant digits). Varying parameters such as input and output activity density lowered the performance of all networks since a less unique activation was possible, but their relationship with HC remained similar, i.e., Random, Axons-Choose, and Dendrites-Choose perform worse than HC under these alternative parameters as well. These results support the argument that the relationship between performance and network parameters can be complex.

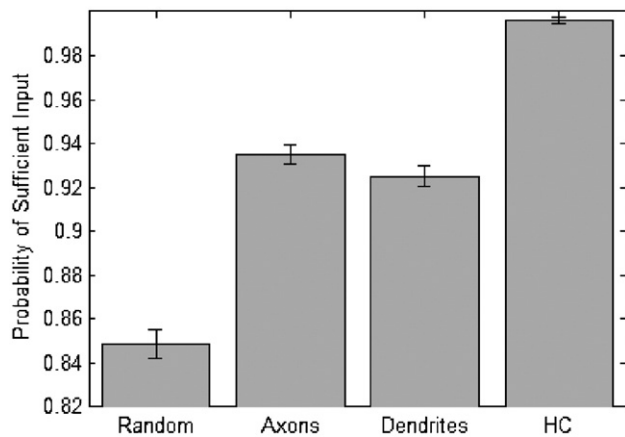


Fig. 3 – Probability that sufficient input is provided to each of the winning neurons in a layer. Of 100 input neurons, 50 were active per test. The twenty output neurons were activated as a kWTA layer, where $k=10$. Inputs that caused one of the 10 winners to receive less than 5 active inputs were called failures, otherwise sufficient input is considered to be provided to each active output neuron. The probability of sufficient input is shown (mean ± S.E.) for the four synaptic distribution classes. HC attained almost perfect performance, whereas the random scheme provided the worst (16% of inputs caused failures), with Axons-Choose and Dendrites-Choose intermediate, at ~8% failures. Two thousand samples of each network were taken to estimate the probabilities.

Random and Axons-Choose distributions yield networks that exhibit substantial information loss. Dendrites-Choose methods lose some information, and hypergeometric connectivity patterns exhibit no information loss. It is notable that, in this task, the hypergeometric distribution of synapses enables sparsely connected networks to perform as well as networks with full dense connectivity.

One reason to expect the decreased performance of Axons-Choose and Random is that these two connectivity schemes lead to output neurons with more or fewer inputs than others. This can lead to particular neurons winning more or less often than others. The extreme case, in which one neuron is always active, can be seen intuitively to reduce the number of unique outputs from 20-choose- k $\binom{20}{k}$ to 19-choose- $(k-1)$ $\binom{19}{k-1}$, yielding an increased likelihood of two inputs to cause the same output.

2.2. Experiment 2: probability of sufficient input and failure rate

The second metric was formulated by considering the error conditions of competitive networks. For many computations, a certain minimum amount of input is required to properly carry out a function. An example of this is a neuron performing the operation of “any 3 of 5”, which is a function that specifies that for a particular set of 5 inputs, if any 3 are active then the neuron should fire, or at least compete to fire. When fewer than three inputs are active, it should not fire. In competitive networks, competition, rather than individual

neuron threshold, limits the number of firing neurons. This sometimes results in neurons firing that did not receive very much input, and the result is less accuracy in the network output. When less than this minimum amount of input is present to the k best matching neurons, a failure is said to occur, and an analysis of the frequency of insufficient input can be made (called failure rate). Fig. 3 depicts the finding that networks with hypergeometric connectivity are most capable of providing a certain minimum amount of input to the k neurons that will eventually fire (see Section 4.2 for details). Axons-Choose and Dendrites-Choose perform similarly to each other, while the performance of the Random pattern was significantly lower than all other patterns.

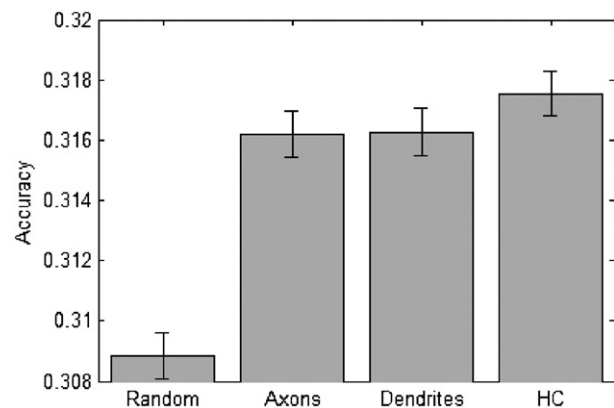


Fig. 4 – Performance of networks at one-step-look ahead part-of-speech tagging task. Results shown (mean ± S.E.) are averaged over 6400 networks per connectivity pattern, ranging from 5% to 50% connection density uniformly. The network recurrently activated itself, and, during learning, used the next incoming part-of-speech activity to determine winners with which to strengthen synapses. The training set consisted of 6000 word pairs randomly selected from the spoken portion of the ICE-GB corpus, 300 per preceding part-of-speech. The testing set consisted of the entire 637,562 words of the spoken portion of ICE-GB. Twenty distinct sets of 5 neurons represented the different selections of the part-of-speech input. For each test input part-of-speech, 5 neurons were activated, and then provided recurrent input for the next time step. The recurrent input selected 5 new neurons to activate, based on the LTP that occurred during the training phase. The set of neurons with the highest number of activated members was interpreted as the part-of-speech predicted by the network to be the successor. 6400 networks for each synapse distribution pattern were tested, and the results were averaged. A fully connected Markov Model achieved 37% tagging accuracy, not far above the performance shown by the models described here. Performance improved as the connection density increased (not shown). Hypergeometric connectivity maintained the most robust performance across the range of densities, averaging 31.75% part-of-speech prediction accuracy. Axons-Choose and Dendrites-Choose connectivity patterns averaged 31.6% prediction accuracy, and Random connectivity fared worse, at 30.09% accuracy.

2.3. Experiment 3: a real-world task: part-of-speech tagging

To test whether hypergeometric connectivity networks would show differential performance in a real world learning environment, a task intended to measure the performance of the networks at part-of-speech sentence tagging (D'Este et al., 1999) was implemented. Sentences from the spoken portion of the British component of the International Corpus of English (Nelson, 1998) were used to provide sequences of parts-of-speech. Twenty basic parts of speech were present, and activation patterns were assigned to each. The recurrent neural network used a very simple LTP learning rule that slightly incremented weights between previously and currently active neurons up to a biologically sound ceiling of 2.0 (Kilborn et al., 1996; Rodriguez et al., 2004). The results (Fig. 4) reinforce the findings of the information maintenance and sufficient input tests, showing an advantage for the hypergeometric distribution pattern (see Section 4.3 for details).

Different from the other two experiments, however, is the fact that the Axons-Choose and Dendrites-Choose patterns appear to perform not just similarly, but exactly the same. This could suggest that the key factor for this metric is the likelihood of a connection existing at all between two neuron groups (representing two parts of speech). Since connection density was varied from 5% to 50%, the 5% case provides the scenario where each set of 5 neurons connects with up to 25 other neurons, making it unlikely that each unique set of 5 neurons (representing each of the 20 parts of speech) contains a connection to each other set. When such a connection is lacking, it is not possible for certain parts-of-speech to be

“guessed” following certain other parts of speech. Due to the recurrent nature of the networks in this experiment, the probability of such a connection existing is the same in the Axons-Choose and Dendrites-Choose, which may cause identical performance.

2.4. Experiment 4: length of nonrepeating sequences generated

The fourth metric, derived from Minai and Levy (1993), measured the length of firing pattern cycles. A simple recurrent neural network was made to activate itself continually until a duplicate activation occurred (indicating a cycle). Networks of up to 120 neurons were tested, and preliminary studies with dense connectivity yielded non-recurrent sequences of 20 million iterations. The recurrent network activation loops resemble the “absorbing sets” that have been studied in Markov processes (Chen and Tweedie, 2000). Cycle length before network learning provides an upper bound on the length of sequences that can be learned. Cycle length detection can also be considered an extension of the information maintenance metric since many inputs are collected and output collisions are measured similarly. Thus, long cycles indicate less information loss. The number of unique firing patterns output before cycle detection is shown in Fig. 5 (see Section 4.4 for details).

In order of performance, hypergeometric performs best, maintaining improved performance against over all other connectivity patterns at all network sizes. Dendrites-Choose performs below hypergeometric, followed by Axons-Choose and then Random. Indeed this ordering of performance is

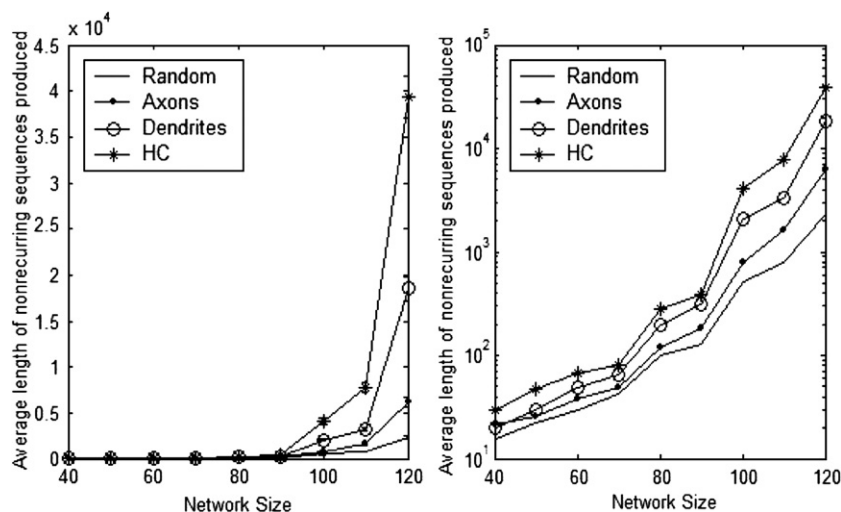


Fig. 5 – Ability to generate long nonrepeating sequences as a function of network size. (Left) Averages of length of sequences produced by recurrent networks with network sizes ranging from 40 to 120, from each of the four network classes. Error bars are shown and are small. A network’s output at time $t=0$ provided input to the network at time $t=1$. A hash table was used to store the activation patterns so that duplicate activations could be detected. Shown on the y-axis is the number of activation cycles that occurred before a duplicate activation was detected. The input density (number of active input/output neurons) was held to 10%. Larger networks have the ability to represent more possible activation patterns, so recurrent cycle length increases with network size. One hundred networks were generated per combination of network parameters to obtain more robust averages. Random and Axons-Choose networks did not generate long cycles. As network size grows, hypergeometric connectivity (HC) begins to show a non-linear improvement over Dendrites-Choose networks. (Right) Study on the log-scale shows that the gap in performance is maintained or increases as an exponential improvement.

maintained for all network sizes except for a single special case in which Axons-Choose performed better Dendrites-Choose in the smallest network examined (40 neurons).

2.5. Experiment 5: scaling: effect of network growth

It is notable that, in the previous experiment, two classes of network, Random and Axons-Choose, fail to yield long cycles (see Fig. 5). The two top performers, Dendrites-Choose and hypergeometric connectivity (HC), were investigated further. These patterns were analyzed using the failure rate metric, and brain growth was simulated by increasing the number of neurons across tests. Fig. 6 shows that as the network size grows, Dendrites-Choose suffers performance loss but HC is able to maintain nearly perfect performance (see Section 4.5 for details). Thus, HC appears uniquely able to scale with network size in this task.

One reason to expect the decreasing performance of Dendrites-Choose can be seen when analyzing the change from growth factor 1 to 2. At growth factor 1, the probability of

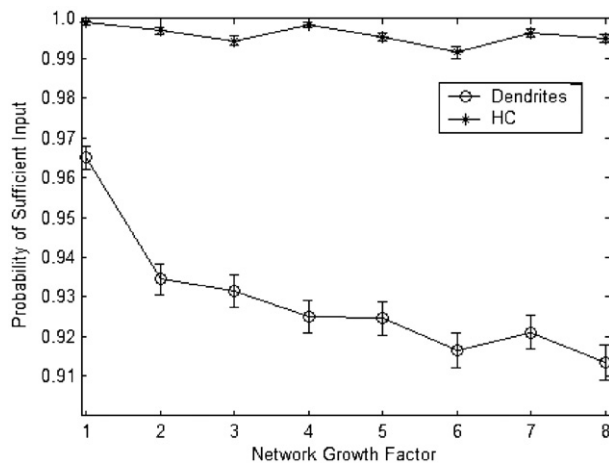


Fig. 6 – Probability of sufficient input to drive a network, as a function of network size. Feedforward networks were built with 10 output neurons. The number of input neurons varies as the x-axis from 10 (growth factor 1) to 80 (growth factor 8). Connectivity was held constant at 30%. Dendrites-Choose and HC were studied with 100 network samples each per growth factor. As in Fig. 3, networks were tested for their ability to avoid “failures” associated with being driven by insufficient inputs (see text). In this scenario, 3 active input neurons must provide input to at least 5 output neurons. The y-axis is the probability that at least 5 neurons received at least one active input. At the far left, the generated networks have 10 input neurons, 10 output neurons, 3 active input neurons, and 30% connectivity (30 synapses). At that point (growth factor 1), HC provided some input to at least 5 output neurons in almost 100% of all samples (1998 of 2000 samples). In contrast, Dendrites-Choose networks with the same parameters provided input to at least 5 output neurons in only 96.5% of samples (1930 of 2000 samples). Once the input layer is increased to 80 neurons (240 synapses total), the figure is still virtually perfect for HC (1990 of 2000 samples) but Dendrites-Choose drops to 91% (1820 of 2000 samples).

an arbitrary output neuron receiving input from an active input neuron is 71%. At growth factor 2, however, this drops to 68%. With respect to hypergeometric, one performance-improving effect could be that if an output neuron connects to a particular input neuron that does not receive active input, then it is known that another output neuron is less likely to connect to that same inactive input neuron, since one of that input neuron’s synapses has been used up (and it will connect to only two other output neurons). Thus, in the hypergeometric case, if one output neuron’s synapse does not receive active input, then the probability of another output neuron receiving active input is increased. The actual probabilities operating over the entire experiment are more complicated, but this example provides a simple illustration of two effects that may be occurring in the performance observed in this experiment.

3. Discussion

The results show that networks connected via hypergeometric distributions exhibit both quantitative and qualitative performance superiority over the other three classes of network circuit design. Random and Axons-Choose patterns often performed far below Dendrites-Choose and hypergeometric connectivity. In particular, the former two classes showed a fundamental inability to avoid repeating activity cycles in recurrent network tests, rendering them unable to generate long sequences (in experiment 4). Several experiments were key illuminators of the differences between the remaining two network types, Dendrites-Choose and hypergeometric connectivity. As the networks grow larger, hypergeometric connectivity exhibits a widening advantage over Dendrites-Choose, suggesting that hypergeometric connectivity may play a role in the crucial ability of mammalian brain circuitry to retain functionality, let alone to improve it, while scaling from milligrams of weight and tens of millions of neurons (e.g., in rodents) to kilograms and tens of billions of neurons (in large-brained mammals such as humans) (Granger, 2006).

A constructive understanding of the principles underlying the observed results may be gained by considering the simple task illustrated in Fig. 7. Starting from a hypergeometrically connected network (upper left), incremental connectivity changes are made to convert the network to a Dendrites-Choose (lower left), Axons-Choose (upper right), or Random (lower right) network. Each sample network has four input cells (above) projecting to four output cells (below), with the synapses from input to output illustrated by connecting lines.

The performance of each network is then depicted twice, illustrating its responses to two different input activity patterns. In each instance, the left-hand network has cells 1 and 2 active (darkened), whereas the right-hand network has cells 1 and 4 active. Active input cells (above) send signals (via darkened axons) to target cells (below); the target cells that are most activated are selectively darkened, illustrating the effect of lateral inhibition (as previously described).

In the hypergeometric net, each input cell (above) is connected via a fixed number of contacts (in this case two) with each target cell (below). On the left, activating cells 1 and 2 provide the most activation to target cell 2, whereas on the

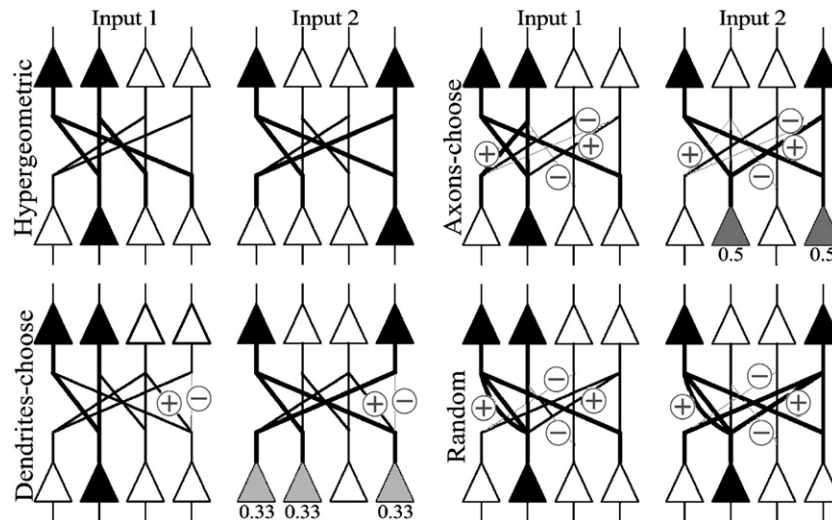


Fig. 7 – Principles underlying hypergeometric superiority across multiple tasks. Shown are four networks, each responding to two different input patterns (see text). For each network, input layers (above) project via synaptic connections (lines) to target cell layers (below). The upper left diagram depicts a hypergeometrically connected circuit. All other circuits are constructed via excising (\ominus) and inserting (\oplus) different connections, to create Dendrites-Choose (lower left), Axons-Choose (upper right), and Random (lower right) networks. As described in the text, the hypergeometric circuit effectively discriminates between two different similar input activity patterns (cells 1 and 2 versus cells 1 and 4); all other circuits discriminate less well or not at all, exhibiting collisions between the output responses to the two different input patterns. It can be seen why this general principle obtains throughout networks of these types, underlying the plethora of observed results, in which hypergeometric networks uniquely and consistently outperform other networks regardless of task.

right, activating cells 1 and 4 provide most activation to target cell 4. Thus, two different input patterns elicit two different responses.

The “Dendrites-Choose” diagram (lower left) was created by eliminating one connection (\ominus) in the hypergeometric net, and adding (\oplus) a different connection, rendering a Dendrites-Choose net (i.e., one in which target cell dendrites all have an equal fixed number of synaptic inputs but input cells do not necessarily generate the same number of synapses). In this case, input layer activity in cells 1 and 2 still yields target cell 2, but input cells 1 and 4 now leads to a tie between cells 1, 2, and 4—one of which (cell 2) is the same cell that was active in response to the previous input (input cells 1 and 2), i.e., a response collision between two different input patterns.

The third diagram was created by again starting with the hypergeometric net, and now eliminating (\ominus) two different axons from the input layer and substituting two new added (\oplus) connections, yielding an Axons-Choose network (fixed number of axons produced from input layer, variable number of synapses per dendrite in target layer). Again, activity in cells 1 and 4 yields a tie among target cells, and one of the activated targets again collides with the activation (cell 2) elicited by a different input pattern (cells 1 and 2).

Finally, the fourth diagram was created from the hypergeometric net by eliminating two of its connects and substituting two others. In this case, both input patterns (cells 1–2 and cells 1–4) activate the same target cell (cell 2), again yielding a collision of responses to different inputs, failing to distinguish the distinct input patterns.

The results still just scratch the surface of understanding the specific relationships between each pair of connectivity

pattern, addressing important issues but also raising new subtle questions (for example, why does Dendrites-Choose perform better than Axons-Choose at certain tasks under certain parameters, but not others?), which are not the focus of the present paper. The results here clearly show that hypergeometric connectivity yields circuits that outperform differently connected networks, as well as unveiling additional subtle yet powerful effects of connectivity patterns, that until now have not been the subject of much discussion or systematic investigation.

While the present study has focused on anatomically stable synaptic distributions of the kind that are implicated in adult mammalian brain circuits, other research has focused on possible effects of neural pruning after learning (Chechik et al., 1999; Mimura et al., 2003; Iglesias et al., 2005). Such work differs in many ways; notably in that it measures performance as the ability to avoid memory loss rather than the ability preserve information and provide a fertile network for new learning.

We chose four types of task to test the functional capabilities of the different classes of networks, but there are of course many other tasks that could be chosen. Since neural network models found in the literature use widely varying learning rules, improved generality of the findings was sought by including experiments in which synaptic weights were not modified. The findings suggest the utility of future work focused on performance measures involving additional learning scenarios. One particular test, not performed here, might provide an interesting subject for future study. Blind source separation is the task in which a single input stream that combines sounds from several sources (e.g., guitars, speech,

and automobiles) is to be separated into appropriate separate sound streams. Under appropriate circumstances, this task can be performed by approaches including Independent Component Analysis (ICA), and non-negative sparse matrix multiplication (NSM) (Cichocki and Georgiev, 2003). Testing an ICA unmixing matrix created with the four different possible sparse synaptic distribution patterns studied here would provide an additional, real-world task that may be performed similarly in actual brain computation. It might seem that using hypergeometric distributions would place undue extra constraints on an ICA unmixing matrix, but we conjecture that the added constraint might provide a means of accelerating the algorithms that derive the unmixing matrix, by limiting the search space.

It is interesting to note the resemblance between the hypergeometric connectivity patterns studied here, and the null hypothesis matrices used in data mining and biogeographical ecology to ignore spurious findings (Gotelli and Entsminger, 2003; Gionis et al., 2006). The question of whether naive neural networks can be used to determine the level of confidence to place in the output of learned networks is an intriguing future area of research to pursue.

More advanced models may require the use of multiple connectivity models to approximate the effects of input from multiple different cortical or subcortical layers or regions. It may also be that the apical and basal dendrites of pyramidal neurons create different random synapse distributions, and the effects of using multiple distributions provide a subject for future study.

Some recent data are notable for their relevance to the hypotheses forwarded here. For instance, spine density within a given dendritic arbor on nonpyramidal cells tend to be highly uniform (e.g., Kawaguchi et al., 2006), as would be expected in hypergeometric circuits (and in Dendrites-Choose, but not in Axons-Choose or Random arrangements). In addition, superficial layer pyramidal cells projecting from one cortical region to another show laminar differences (e.g., Lund et al., 1993), and studies of the connections of synaptically coupled superficial layer pyramidal cells exhibit relatively low variance, consistent with the hypergeometric hypothesis.

Of particular note are recent findings by Feldmeyer et al. (2006), studying synaptically coupled cells within superficial layers of a neocortical column. They report data “suggesting that 270 pyramidal cells in layer 2/3 are innervated by an individual pyramidal cell. In turn, a single pyramidal cell received synaptic inputs from 270 other L2/3 pyramidal cells” (Feldmeyer et al., 2006). These exciting recent data appear strongly to support the hypergeometric connectivity hypothesis.

Taken together, the findings here indicate that the choice of distribution pattern used for partially connected neural networks is of substantial significance, capable of changing performance characteristics in unexpected ways. The results provide evidence that hypergeometric connectivity is very unlikely to yield accidental negative results from empirical neural modeling tests, and the range of tasks used here suggests the generality of the findings. We specifically conjecture (i) that neocortical circuits are connected via hypergeometric distributions, or close approximations of these distributions; (ii) that these connectivity patterns confer

some of the computational power of neocortical circuitry; and (iii) that the expensive mechanisms of synaptic overgrowth and subsequent pruning exhibited during brain development may in part occur as a way of arriving at close approximations of hypergeometrically connected brain circuits.

Finally, it is useful to place these findings and conjectures into a larger context. Just as connectivity constraints are shown here to yield benefits in resulting brain circuit computation, additional anatomical and physiological constraints can lead to unexpected, and unexpectedly powerful, algorithms and circuit designs. In particular, the intrinsic parallel nature of the algorithms derived from brain circuits makes them unusually well suited to highly parallel hardware implementations. It has recently been shown that such designs lead to implementations of vision processing systems that equal or outperform extant systems, including performance accuracy measures, speed increases, and reduced power consumption (Felch et al., 2007; Furlong et al., 2007). Ongoing research aims to incorporate such findings to address longstanding issues in computational architecture design, including multiprocessor and cluster communication, and memory latency problems, which otherwise impede the creation of large, real-time brain circuit simulations.

4. Experimental procedures

In neocortical local circuits, the majority of neurons are excitatory and project information to potentially distal targets; whereas a smaller number of inhibitory neurons receive input from and output to only local excitatory neurons. This leads to the behavior of competitive networks, where maximally activated excitatory neurons fire first, resulting in activation of local inhibitory cells that suppress additional neuron firing in all excitatory cells, including the ones that initially fired. When only one neuron activates, it is often termed a “winner-take-all” network, with the responding neuron “winning” the race to fire. When more than one neuron activates before the suppressing wave of inhibition, it is called a “k-winner-take-all” (kWTA) network, where k is the number of winners. Specific instantiations of the family of networks described above were simulated such that existing synapses were

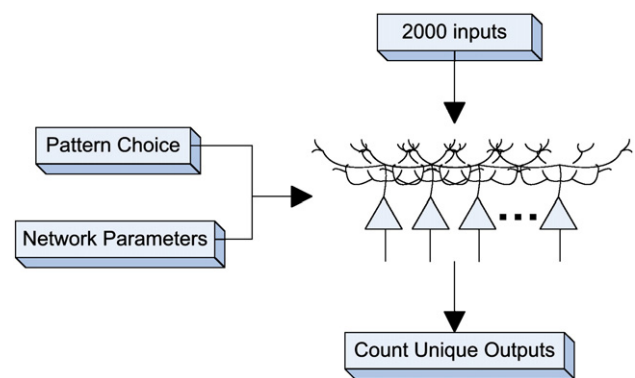


Fig. 8 – Experiment 1. Two thousand unique inputs are presented to a kWTA feedforward network with a specified random connectivity scheme, connection density, and activity densities. The number of unique outputs is counted.

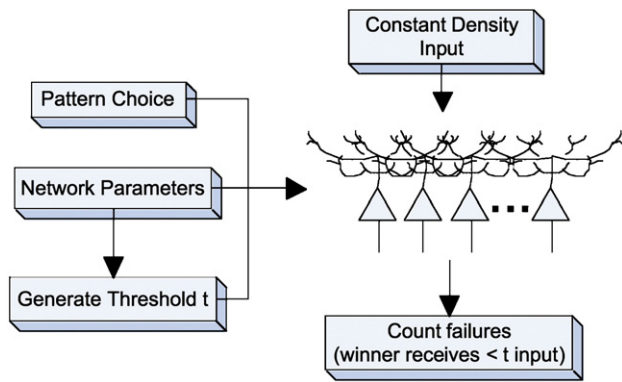


Fig. 9 – Many networks are generated for each of the four connectivity patterns and input activates the kWTA layer. A threshold is selected and the probability that one of the winners receives less than this amount of input is measured.

initialized with random weights close to 1.0 (increasing, via LTP, up to a biologically realistic ceiling of 200%, or 2.0, in experiment 3); and the binary neuron threshold was set to the net excitation of the k -th most active neuron.

4.1. Design and execution of experiment 1

Experiment 1 studies the effects of information maintenance and loss by measuring the probability that two unique inputs “collide” and cause the same output. When this occurs in the presence of learning, it is called clustering; but this experiment used naive networks to test the synapse distribution patterns to preserve information rather than manipulate it. Two layer feedforward networks of each class were prepared with 20 input neurons, 20 output neurons, and 10% connectivity (in this case, 40 synapses). For comparison, full connectivity (100% connection density) was also included in this experiment. The input layer was activated randomly (but uniquely) 2000 times per sample. The input and output activity was held to 50% density (10-active, 10 non-active) and the output network operated as a k -winners-take-all network, with $k=10$. Ties were made unlikely through the slight bumping of naive synapses so that each synapse was very close to a 1.0 weight value. In the event of ties, the network reduced activity so that the number of winners was always less than or equal to 10. One hundred samples were

taken with each network, each sample yielding a count of the unique outputs induced by 2000 unique inputs. Fig. 8 depicts the experimental design and Fig. 2 shows the results.

Dendrites-Choose networks were created by assigning the 40 synapses equally among the 20 output neurons, so that each output neuron received the same number of inputs. In this scenario, input neurons had a variable number of outputs, determined by the random selection made by each output neuron’s dendrites.

Axons-Choose networks were created by assigning the 40 synapses equally among the 20 input neurons, and each output neuron received a variable number of inputs determined by the random selections made by the axons.

Random connectivity iteratively chose an input and output neuron pair, with replacement, over 40 iterations. Thus, it was possible for two neurons to be connected to each other with more than one synapse; a scenario not shared with the other connectivity methods.

Hypergeometric connectivity networks require special care to construct because there exists no perfect heuristic by which synapses can be allocated greedily without backtracking. Brute force attempts to form hypergeometrically connected networks were too cumbersome, and failed completely in larger networks. A method was implemented that untwisted a non-hypergeometric network until it became hypergeometric. First, a network was generated with the Axons-Choose method. Then, output neurons that received too many inputs had a synapse pruned, and the input neuron that shared it (now deficient one synapse) grew a new synapse with an output neuron not already connected to it. This was done iteratively, and is an idealized case of what overgrowth and pruning may carry out in biology.

4.2. Design and execution of experiment 2

Experiment 2 was designed to understand the capabilities of the connectivity methods at providing at least a minimum amount of active input to output neurons that would eventually activate as one of k -winners. The question arose when it was discovered that certain inputs can be presented to Dendrites-Choose networks such that no output neuron receives any active input. This occurs if the active input neurons do not synapse with any output neurons. Studying the problem more generally, 3000 networks for each connectivity scheme were constructed, each with 100 input neurons,

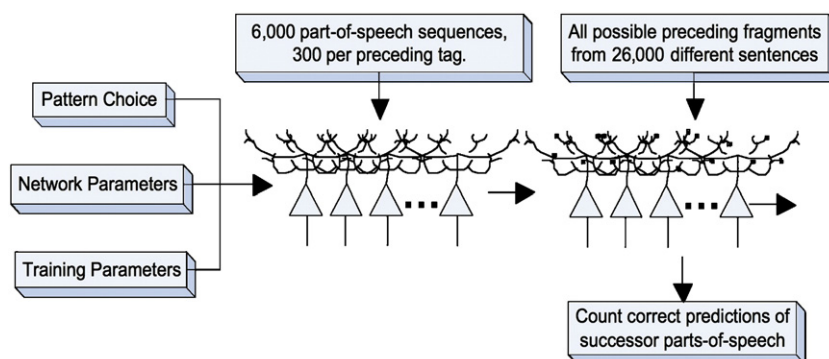


Fig. 10 – Design of experiment 3, yielding part-of-speech tagging accuracy after learning.

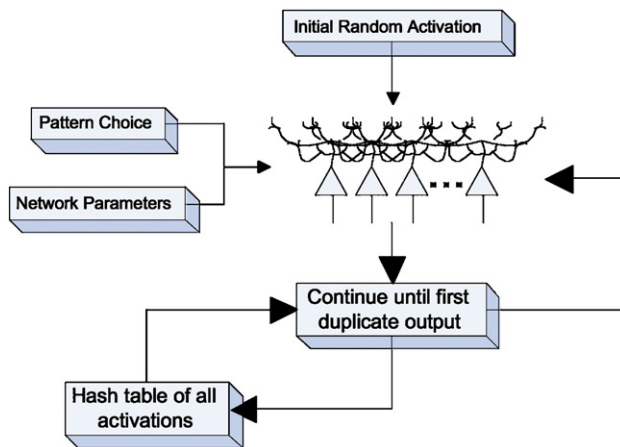


Fig. 11 – Experimental design for estimating the length of non-recurring sequences produced by recurrent activation.

20 output neurons, 50% activity density, and 10% connection density. A measure of the probability that at least one of the 10 winning output neurons received less than five active inputs was recorded. The experimental design is shown in Fig. 9, and the results of the experiment in Fig. 3.

4.3. Design and execution of experiment 3

Experiment 3 was designed to see if the performance effects shown in experiments 1 and 2 would give rise to performance differences in a real world task. Part-of-speech tagging is the process of assigning a label such as “noun” or “verb” to words in input sentences. The spoken portion of the ICE-GB corpus was used, and sequences of the 20 tags assigned by linguists for that corpus were represented as non-overlapping network activity patterns. Networks of 100 neurons were created, five of which were active during any given time step, representing the unique set of five neurons assigned to represent the particular input part-of-speech. Each network was directly recurrent to itself using one of the four connectivity patterns under study, and the training and testing sets were divided into thousands of individual sequences of two parts-of-speech. During both training and testing, the appropriate five neurons activated for the first part-of-speech input. During training, the second input part-of-speech would cause activity in the same way, and the recurrent synapses between the first and second part-of-speech activation patterns were strengthened. During testing, only the recurrent input drove the network during the second time step, and the resulting 5-hot activation pattern was compared with the unique set of neurons assigned to each part-of-speech to interpret the firing pattern as a predicted part-of-speech. During training, the synapses between the neurons activated by the previous word and the current word were strengthened. The training set consisted of 6000 part-of-speech sequence pairs, 300 for each possible preceding tag. For simplification, no synapse saturation prevention mechanism was implemented, so a low learning rate of 0.001 was used to make saturation unlikely. In the high connection density scenarios such as the 50% case, prediction performance for all networks

was above 35%, very close to the 37% accuracy that a Markov model produced while verifying baselines.

During training, the next input word drove the current network activity, whereas during testing the recurrent input of the network feeding into itself drove the activity. The activity was interpreted such that a single part-of-speech was guessed. The accuracy of these guesses was measured for each of the connectivity patterns, 6400 network samples for each of the four patterns, uniformly over a range of 5% to 50% connection density in 5% increments. Fig. 10 depicts the experimental design and Fig. 4 shows the results.

4.4. Design and execution of experiment 4

Experiment 4 was designed to study whether the connectivity pattern effects on information maintenance have implications for the ability of a network to recurrently activate itself without repetition. As before, a kWTA network was recurrently connected to itself using one of the four connectivity patterns. The network was analyzed in various sizes between 40 and 120 neurons at increments of 10. k was set to 10 and the connectivity density was set to 5%. For each network size and connection method, 100 networks were generated, each yielding a performance sample. Each sample counted the number of recurrent activations produced before the first duplicate. Fig. 11 shows the experimental design, and Fig. 5 shows the average length of the non-recurring sequences produced.

4.5. Design and execution of experiment 5

In experiment 5, the Dendrites-Choose and HC connectivity patterns are compared as the size of the network grows. The metric is the same as in experiment 2. Connectivity density was held constant at 30%, and the output layer size was held to 10 neurons. The input layer size varied from 10 to 80 neurons in 10 neuron increments. Three input neurons were active during any given input activity pattern. The result is that input activity density varies from 30% to 3.75%. One hundred network samples were taken for each pattern and network size. The output layer was assigned a k value of 5, and the input threshold used for failure detection was set to 0.5. Thus, a failure occurred if less than five output neurons received input. Fig. 6 shows the results. HC appears to maintain nearly perfect performance, whereas Dendrites-Choose starts with a 3.5% failure rate that increases as the network size grows, doubling to 7% once the input layer has grown to 4 times its original size.

4.6. Generality of experimental designs

All of the experiments could be run in both kWTA and pure threshold versions. The sufficient input metric of experiments 2 and 5 does not require that the network be kWTA and so those results already generalize to arbitrary feedforward networks. The part-of-speech tagging task of experiment 3 would also generalize since the definitions of part-of-speech input used there could be used in non-kWTA networks. Increased range of synapse weight beyond the 100% to 200% implemented here could increasingly allow synapses to simulate their own non-existence (with a weight of 0.0) and this would certainly affect results.

Information maintenance and recurrent cycle experiments would likely yield similar results beyond kWTA since they are affected by the amount of information capable of being output, which in turn is affected by neurons always or never winning, which can still occur in neurons with a static, rather than competitively determined, threshold.

Acknowledgments

This work was supported in part by grants from the Office of Naval Research (ONR) and from the Defense Advanced Research Projects Agency (DARPA).

REFERENCES

- Adelsberger-Mangan, D.M., Levy, W.B., 1992. Information maintenance and statistical dependence reduction in simple neural networks. *Biol. Cybern.* 67, 469–477.
- Amari, S., Arbib, M.A., 1977. Competition and cooperation in neural nets. *Systems Neuroscience*. Academic Press, New York.
- Bannerman, D.M., Good, M.A., Butcher, S.P., Ramsay, M., Morris, R.G.M., 1995. Distinct components of spatial learning revealed by prior training and NMDA receptor blockade. *Nature* 378, 182–186.
- Bell, A.J., Sejnowski, T.J., 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159.
- Bose, J. (2003). A scalable sparse distributed neural memory model. Unpublished Dissertation, University of Manchester.
- Braitenberg, V., Schüz, A., 1998. *Cortex: Statistics and Geometry of Neuronal Connectivity*. Springer, Berlin.
- Brunel, N.G., 2000. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.* 8, 183–208.
- Chechik, G., Meilijson, I., Ruppin, E., 1999. Neuronal regulation: a mechanism for synaptic pruning during brain maturation. *Neural Comput.* 11, 2061–2080.
- Chen, P., Tweedie, R.L., 2000. Orthogonal measures and absorbing sets for Markov chains. *Math. Proc. Camb. Philos. Soc.* 121 (01), 101–113.
- Cichocki, A., Georgiev, P., 2003. Blind source separation algorithms with matrix constraints. *IEICE Trans. Fundam.* E86-A (3), 522–531.
- Coultrip, R., Granger, R., Lynch, G., 1992. A cortical model of winner-take-all competition via lateral inhibition. *Neural Netw.* 5, 47–54.
- D'Este, C., Towsey, M., Diederich, J., 1999. Sparsely-connected recurrent neural networks for natural language learning. *First Workshop on Natural Language Processing and Neural Networks*. Beijing, China, pp. 64–69.
- Elizondao, D., Fiesler, E., Korczak, J., Idiap, M., 1995. Non-ontogenic sparse neural networks. *IEEE Int'l Conf Neural Networks*, pp. 290–295.
- Feldmeyer, D., Lubke, J., Sakmann, B., 2006. Efficacy and connectivity of intracolumnar pairs of layer 2/3 pyramidal cells in the barrel cortex of juvenile rats. *J. Physiol.* 575, 583–602.
- Felch, A., Moorankikara-Nageswaran, J., Furlong, J., Chandrashekar, A., Dutt, N., Nicolau, A., Veidenbaum, A., Granger, R., 2007. Accelerating brain circuit simulations of object recognition with a Sony PlayStation 3. *International Workshop on Innovative Architecture for Future Generation High-Performance Processors and Systems (IWIA 2007)*, IEEE Press.
- Furlong, J., Felch, A., Nageswaran, J., Dutt, N., Nicolau, A., Veidenbaum, A., Chandrashekar, A., Granger, R., 2007. A new class of intrinsically parallel brain-derived algorithms scales linearly with number of processors. *Proceedings of the International Conference on Parallel Computing*, (parco.org) 2007.
- Fiesler, E., 1994. Comparative bibliography of ontogenic neural networks. *Proceedings of the International Conference on Artificial Neural Networks*, pp. 793–796.
- Gionis, A., Mannila, H., Mielikainen, T., Tsaparas, P., 2006. Assessing data mining results via swap randomization. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 167–176.
- Gotelli, N.J., Entsminger, G.L., 2003. Swap algorithms in null model analysis. *Ecology* 84, 532–535.
- Granger, R., 2006. Engines of the brain: the computational instruction set of human cognition. *AI Magazine* 27, 15–32.
- Grossberg, S.K., 1976. Adaptive pattern classification and universal recoding: parallel development and coding of neural feature detectors. *Biol. Cybern.* 23, 121–134.
- Iglesias, J., Eriksson, J., Grize, F., Tomassini, M., Villa, A.E., 2005. Dynamics of pruning in simulated large-scale spiking neural networks. *Biosystems* 79, 11–20.
- Kawaguchi, Y., Karube, F., Kubota, Y., 2006. Dendritic branch typing and spine expression patterns in cortical nonpyramidal cells. *Cereb. Cortex* 16, 696–711.
- Kilbom, K., Lynch, G., Granger, R., 1996. Effects of LTP on response selectivity of simulated cortical neurons. *J. Cogn. Neurosci.* 8, 328–343.
- Lund, J.S., Yoshioka, T., Levitt, J.B., 1993. Comparison of intrinsic connectivity in different areas of macaque monkey cerebral cortex. *Cereb. Cortex* 3, 148–162.
- Mimura, K., Kimoto, T., Okada, M., 2003. Synapse efficiency diverges due to synaptic pruning following overgrowth. *Phys. Rev. E* 68, 31910.
- Minai, A.A., Levy, W.B., 1993. Dynamics of sparse random networks. *Biol. Cybern.* 70, 177–187.
- Morris, R.G.M., Anderson, E., Lynch, G.S., Baudry, M., 1986. Selective impairment of learning and blockade of long-term potentiation by an N-methyl-D-aspartate receptor antagonist, AP 5. *Nature* 319, 774–776.
- Nelson, G., 1998. *The International Corpus of English. The British Component. ICE-GB. ICE-GB. Getting Started*. London: Survey of English Usage. University College London.
- Oppenheim, R.W., 1991. Cell death during development of the nervous system. *Annu. Rev. Neurosci.* 14 (1), 453–501.
- Oppenheim, R., Qin-Wei, Y., Prevette, D., Yan, Q., 1992. Brain-derived neurotrophic factor rescues developing avian motoneurons from cell death. *Nature* 360, 755–757.
- Purves, D., Lichtman, J.W., 1980. Elimination of synapses in the developing nervous system. *Science* 210, 153–157.
- Riout-Pedotti, M.S., Friedman, D., Donoghue, J.P., 2000. Learning-induced LTP in neocortex. *Science* 290, 533–536.
- Rodriguez, A., Whitson, J., Granger, R., 2004. Derivation and analysis of basic computational operations of thalamocortical circuits. *J. Cogn. Neurosci.* 16, 856–877.
- Saucier, D., Cain, D.P., 1995. Spatial learning without NMDA receptor-dependent long-term potentiation. *Nature* 378, 186–189.
- Schutze, H., 1993. Part-of-speech induction from scratch. *Proceedings of the 31st Conference on Association for Computational Linguistics*, pp. 251–258.
- Thomson, A.M., Bannister, A.P., 2003. Interlaminar connections in the neocortex. *Cereb. Cortex* 13, 5–14.
- Yuille, A.L., Grzywacz, N.M., 1989. A winner-take-all mechanism based on presynaptic inhibition feedback. *Neural Comput.* 1, 334–347.