How do neurons operate on sparse distributed representations? A mathematical theory of sparsity, neurons and active dendrites[1]

Subutai Ahmad[1]*, Jeff Hawkins[1]

[1]Numenta, Inc., Redwood City, CA, USA

* Correspondence: Subutai Ahmad, Numenta Inc., 791 Middlefield Road, Redwood City, CA 94063 USA.
sahmad@numenta.com

Keywords: sparse coding, neurons, active dendrites, neocortex, NMDA spike.

Abstract

We propose a formal mathematical model for sparse representations in neocortex based on a neuron model and associated operations. The design of our model neuron is inspired by recent experimental findings on active dendritic processing and NMDA spikes in pyramidal neurons. We derive a number of scaling laws that characterize the accuracy of such neurons in detecting activation patterns in a neuronal population under adverse conditions. We introduce the union property which shows that synapses for multiple patterns can be randomly mixed together within a segment and still lead to highly accurate recognition. We describe simulation results that provide overall insight into sparse representations as well as two primary results. First we show that pattern recognition by a neuron can be extremely accurate and robust with high dimensional sparse inputs even when using a tiny number of synapses to recognize large patterns. Second, equations representing recognition accuracy of a dendrite predict optimal NMDA spiking thresholds under a generous set of assumptions. The prediction tightly matches NMDA spiking thresholds measured in the literature. Our model neuron matches many of the known properties of pyramidal neurons. As such the theory provides a unified and practical mathematical framework for understanding the benefits and limits of sparse representations in cortical networks.

1.   Introduction

A wealth of empirical evidence suggests the neocortex represents information using sparse distributed patterns of activity (Barth and Poulet, 2012). Representations are sparse because at any point in time only a small percentage of neurons are active while the rest are inactive. Representations are distributed because although each active neuron contributes information, it is the set of active neurons that determine what is being represented. The diversity of sparse distributed representations (SDRs) in neocortex is remarkable. Sparse representations exist in early auditory, visual and somatosensory areas (Hromádka et al., 2008; Weliky et al., 2003; Vinje and Gallant, 2000; Crochet et al., 2011). These representations correspond directly to sensory features such as visual edges and audio frequency bands. Representations in higher cortical areas are more abstract and categorical in nature. These areas can encode concepts that have no direct correlation to sensory features (Kiani et al., 2007). Primary motor areas encode sparse motor maps that correspond to specific movements (Graziano and Aflalo, 2007). Premotor areas encode more abstract behavioral

---

"plans" such as the notion of "grasping and inserting food into your mouth" (Graziano et al., 2002). The ubiquity of sparse distributed representations suggests that they enable operations essential for brain function and neural computation. To function effectively these representations must have tremendous capacity and must be extremely tolerant to noise. The exact laws governing their behavior are unknown.

Our experimental understanding of the mechanics of how neurons operate on their inputs has evolved significantly over the last 15 years. The majority of computational models (starting with the McCulloch–Pitts neuron (McCulloch and Pitts, 1943)) assume a linear integration of inputs followed by a non-linear transfer function (Gazzaniga, 2004) (Figure 1A). Experimentally it is now well established that biological neurons are significantly more complex, particularly with respect to synaptic integration. Excitatory neurons, specifically pyramidal cells, have thousands of synapses (Spruston, 2008) and a complex dendritic structure (Figure 1B). Proximal synapses, those closest to the cell body, have a relatively large effect on the likelihood of a cell generating an action potential. However, a majority of the synapses are distal, or far from the cell body. The activation of a single distal synapse has little effect at the soma, but the coincident activation of a cluster of spatially localized synapses generates a large regenerative NMDA spike (Spruston, 2008; Larkum et al., 2009; Antic et al., 2010). NMDA spikes depolarize the cell; thus spatially localized dendritic segments act as coincidence detectors and provide a means for the many distal synapses to play a significant role in the cell's activity (Losonczy et al., 2008). The characteristics of such "active dendrites" has been studied extensively and it is generally understood that pyramidal cells contain a large number of non-linear active dendritic segments that independently modulate cell responses (Branco and Häusser, 2011; Major et al., 2013).

Sparse representations are ubiquitous in neocortex, and the basic properties of active dendrites and pyramidal cells are also thought to be consistent throughout neocortex (Spruston, 2008). It behooves us to ask whether there exist a common set of governing principles related to active dendrites and sparsity that are universal and independent of modality. A number of studies have explored sparse representations from a theoretical perspective. Early work on sparse coding (Olshausen and Field, 1997) suggests that sparse representations of the type found in V1 might satisfy information theoretic optimality criteria. Additional analysis (Kanerva, 1988; Olshausen and Field, 2004) suggests that sparse representations may be particularly convenient for learning and memory systems. However none of these studies have considered realistic neuron models with active dendrites. With a few notable exceptions (Poirazi and Mel, 2001; Legenstein and Maass, 2011) theoretical studies have ignored active dendrites altogether. The studies that do model active dendrites have not explicitly incorporated sparse representations.

In this paper we develop a novel mathematical theory of sparse distributed representations and networks of neurons with active dendrites. We develop our analysis first from the perspective of a single dendritic segment and then from the perspective of a population of neurons. We propose that neurons form just a few synapses to a sparse subsample of active cells in activity patterns they need to recognize. In other words, a small set of spatially localized synapses on a dendritic branch can detect a prototypical pattern in a large population of active cells. We derive theoretical results that characterize the learning capacity and robustness of such representations in the context of neural tissue. The analysis demonstrates that individual neurons have the capacity to learn and classify an enormous number of patterns under extremely unreliable and noisy conditions. The results hold even when they form synapses with a small subset of the cells in the pattern they want to recognize. Additionally, computer simulations of the equations provide some detailed numerical insights, such as natural bounds on the sparsity of representations, and the specific number of synapses required to

initiate dendritic spikes. The results suggest that, due to active dendrites, neurons can operate on sparse distributed representations in a highly robust and efficient manner.

2.   Materials and Methods
2.1.   Model Neuron

We first develop an abstract neuron model and discuss how such neurons might operate on SDR's. Our model departs from the traditional integrate and fire neurons used in artificial neural network models (Figure 1A). Instead we model the neuron as a set of independent dendritic segments, each detecting one or more patterns of activity in some presynaptic area. This is similar in spirit to the models adopted by (Poirazi and Mel, 2000; Losonczy et al., 2008; Larkum et al., 2009; Wu and Mel, 2009). Note that in real neurons synaptic input can be classified by where the dendrite segment is on the cell (proximal, distal basal, and apical) and where the afferents that connect to the dendrite
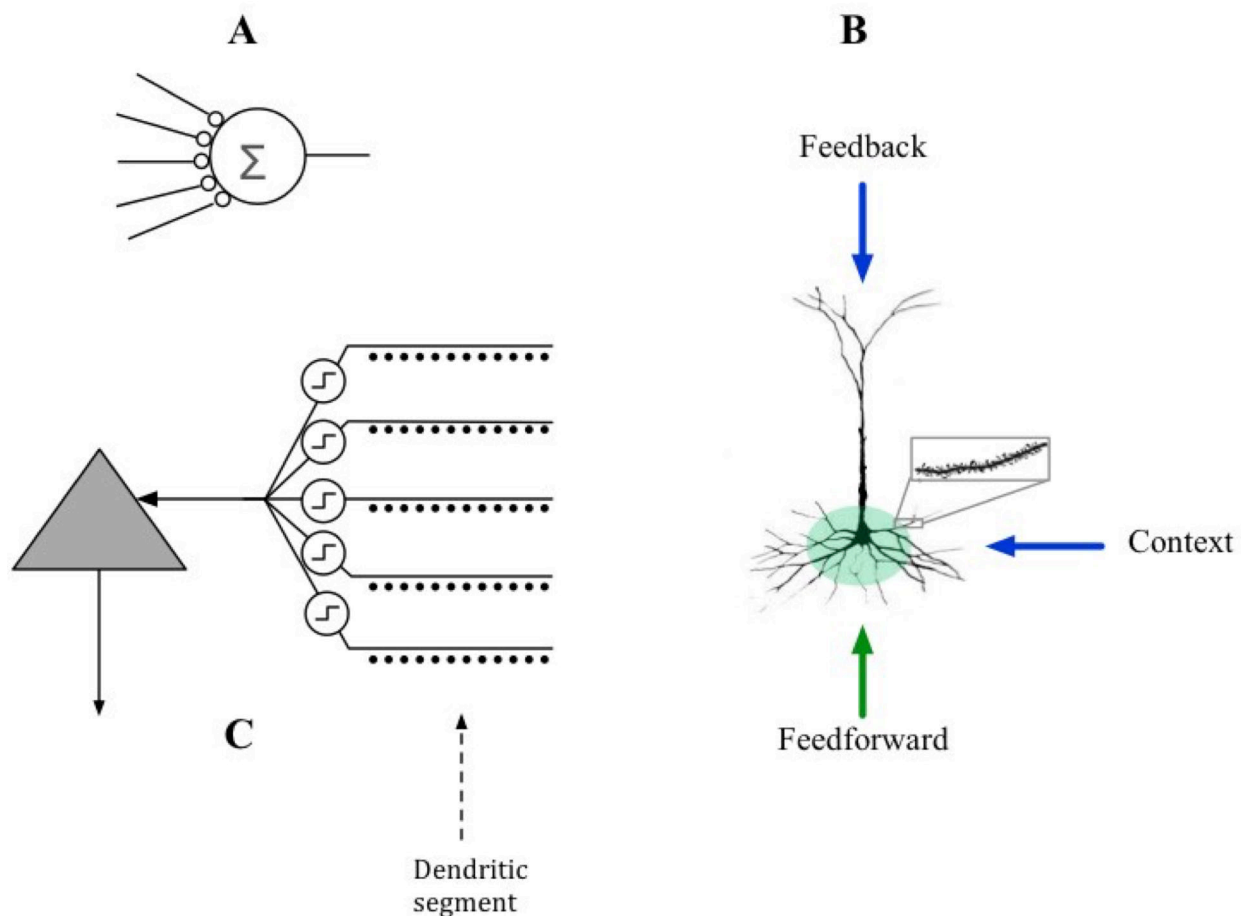

Figure 1. **A**. A point neuron used in most artificial neural networks. The output of the point neuron is a function of the sum of all the inputs; active dendritic properties are ignored. **B.** A neocortical pyramidal neuron has thousands of excitatory synapses that are located on dendrites. Synapses proximal to the cell body (green area) comprise <10% of a cell's inputs, receive feedforward input, and define the basic receptive field response of the cell. Synapses on the basal distal dendrites typically receive contextual input from nearby cells. Apical distal synapses typically receive feedback inputs. The activation of a single distal synapse often has no measurable effect at the soma. However, the activation of a small number of synapses in close spatial and temporal proximity on a basal distal dendrite can cause an NMDA spike and significant depolarization at the soma. **C.** The neuron model we use in this analysis is composed of an array of active dendritic segments (only five shown). Each dendritic segment contains a number of synapses and is associated with a spiking threshold. If the instantaneous number of active synapses on a segment reaches threshold, the segment initiates a spike, thus acting as a coincidence detector.

segment originate. These dendritic zones are shown in Figure 1B. For the present analysis we analyze the recognition capability of generic dendritic segments, as shown in Figure 1C, without regard to the functionality or structure of presynaptic areas. For this paper the segment is generic and agnostic – it simply receives some input and decides whether or not to initiate a dendritic spike.

2.2.   Computing A Match

Figure 2 illustrates our segment model and how a segment detects patterns. We model the instantaneous activity of presynaptic cells as being either on or off. The effect of an individual synapse is similarly binary. As shown in Figure 2A, a dendritic segment would typically be connected to a very small subset of all possible neurons in an input region. The specific subset can change over time as a result of learning (Chklovskii et al., 2004). Therefore the number of *potential connections* to a segment is much larger than the number of actual connections. The actual connections are a subset of prototype activity patterns to be recognized.

Formally, we denote the number of potential connections as n. A dendritic segment is represented as a binary vector $\boldsymbol{D} = [b_0, \cdots, b_{n-1}]$ where a non-zero value b_i indicates a synaptic connection to cell i and $s = |\boldsymbol{D}|$ indicates the number of synapses on that segment. Experimental findings suggest that typical numbers for s are between 20 and 300 (Major et al., 2013). n, the number of potential connections is assumed to be much larger (numbering in the thousands) leading to very sparse $\boldsymbol{D}$.

Similarly, the binary vector $\boldsymbol{A}_t$ represents presynaptic activity (the activity of all potential connections) at time t. $\boldsymbol{A}_t$ is of dimensionality n with $a_t = |\boldsymbol{A}_t|$ as the number of active cells. The activity is assumed to be sparse, i.e. $a_t \ll n$. Figure 2B illustrates such a scenario. When a threshold θ of the synapses on s fire simultaneously, this is a "match" and causes a dendritic spike:

$$match(\boldsymbol{A}_t, \boldsymbol{D}) \equiv overlap(\boldsymbol{A}_t, \boldsymbol{D}) \geq \theta \qquad (1)$$

The overlap score is simply the number of bits that are "1" in the same locations in both vectors (i.e. the dot product):

$$overlap(\boldsymbol{A}_t, \boldsymbol{D}) \equiv \boldsymbol{A}_t \cdot \boldsymbol{D} \qquad (2)$$

The use of the overlap metric with very sparse vectors models the biology of active dendrites and is similar to the model in (Wu and Mel, 2009). It is a departure from most other computational models that typically assume full connectivity between a cell and its input area. An important property of our model is that activity in the presynaptic area that does not correspond to a synaptic connection on a segment has no impact on that segment.
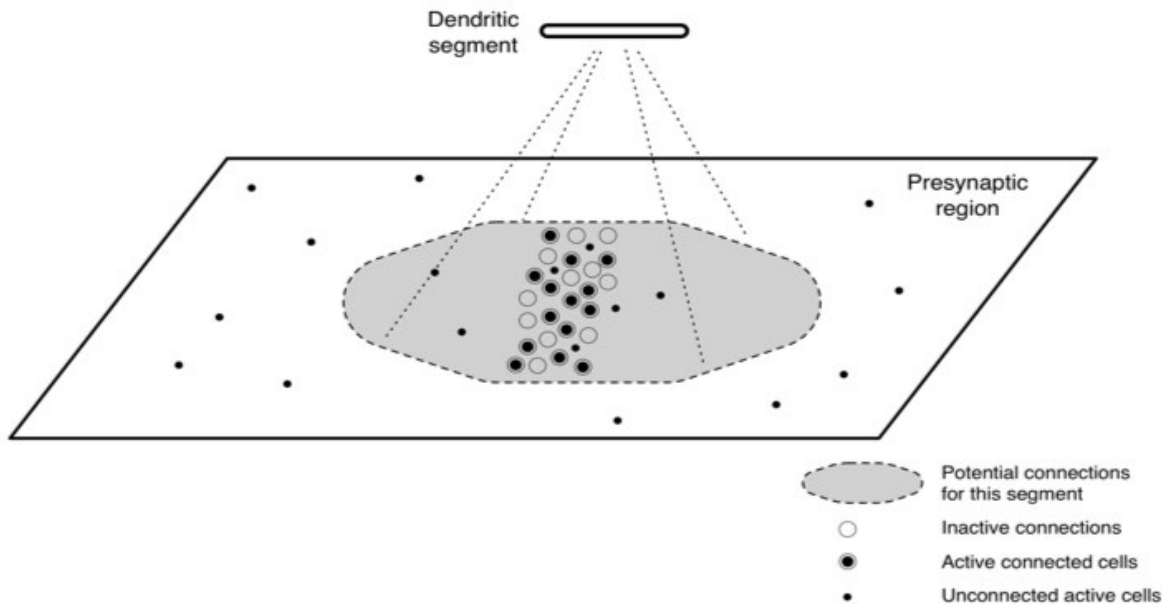
A
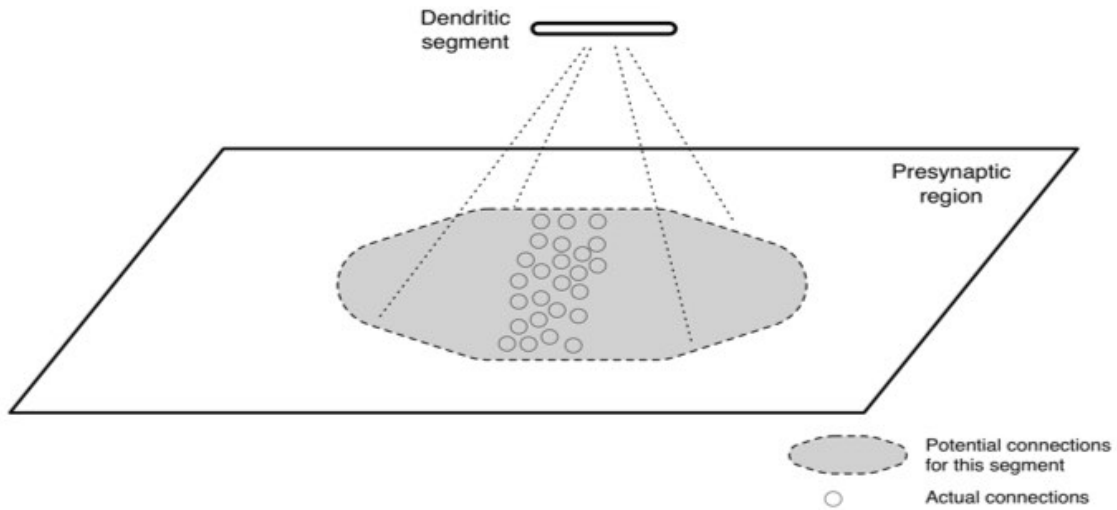
B

Figure 2. **A.** This figure shows a single example dendritic segment and its connections to some presynaptic area A_t. The grey area represents the set of potential connections to this segment. The clear circles denote 25 presynaptic cells corresponding to 25 synapses on the segment. In this example the synapses detect a roughly vertical bar of activity near the center of A_t. The bar is purely for illustration purposes - in general the segment could contain an arbitrary random subsample of A_t. **B.** This figure illustrates the behavior of the segment given a specific pattern of active presynaptic neurons. The small black circles represent the currently active cells in A_t, and filled circles denote synaptic connections that overlap with active cells. In this example there is a fair amount of noise. Many of the cells within the segment's preferred vertical pattern did not become active and there are several additional active cells. However 14 of the active cells are connected; a threshold of 14 or lower will cause the segment to initiate a dendritic spike. When the dimensionality is high, the chance that a random pattern of activity will overlap sufficiently with the segment is extremely low, hence it is very likely that the activity represents an actual vertical pattern, and not an accident.

3.   Results

3.1.   Accuracy of detecting a single pattern on a single segment

The synapses on segments represent a small subsample of the full pattern of activity in presynaptic areas. Given this, how reliably can a single segment detect the larger pattern? We assume that through learning the segment has created synapses to a subset of the active cells corresponding to a given pattern. The segment thus represents some prototypical pattern. We consider the problem of robustly detecting repeat occurrences of that pattern under various distortions. The degree to which a presynaptic pattern has to match the synapses is controlled by θ, the NMDA spike threshold. The lower the value of θ with respect to s, the more deviation or noise the segment can tolerate. The lower the value of θ the more likely the segment would falsely detect a match to a different pattern. Thus there are inherent tradeoffs between the number of synapses per segment, the segment threshold, and the number of potential connections that affect overall noise robustness and the possibility of false matches. This is illustrated conceptually in Figure 3.

In order to quantify this effect, we introduce the notion of an "overlap set". Let X be a binary SDR vector with n components. The overlap set of X with respect to b, $\Omega_X(n, a, b)$, is defined as the set of vectors of size n with a bits on that have exactly b bits of overlap with X. The cardinality of this set, $|\Omega_X(n, a, b)|$, is the number of such vectors. Assuming $b \leq |X|$ and $b \leq a$, $|\Omega_X(n, a, b)|$ can be computed as the number of subsets of X with b bits ON, multiplied by the number of other patterns containing $n - |X|$ bits, of which $a - b$ bits are ON. In other words, this is:

$$|\Omega_X(n, a, b)| = \binom{|X|}{b} \times \binom{n - |X|}{a - b} \tag{3}$$

We can now answer the following question: given a dendritic segment D and random presynaptic activity pattern A_t, what is the probability of a match? In other words, what is $P(overlap(A_t, D) \geq \theta)$? This is simply the number of possible matching vectors divided by the number of possible patterns:

$$P(overlap(A_t, D) \geq \theta) = \frac{\sum_{b=\theta}^{s} |\Omega_D(n, a_t, b)|}{\binom{n}{a_t}} \tag{4}$$

It may be difficult to get an intuitive sense for this equation but as n, s and θ increase, the denominator increases exponentially faster than the numerator. This means that the "grey area" representing the overall space in Figure 3 increases much faster than any of the "white areas" representing possible matches. In other words it is easy to increase the parameters such that the chance of false matches is extremely small.
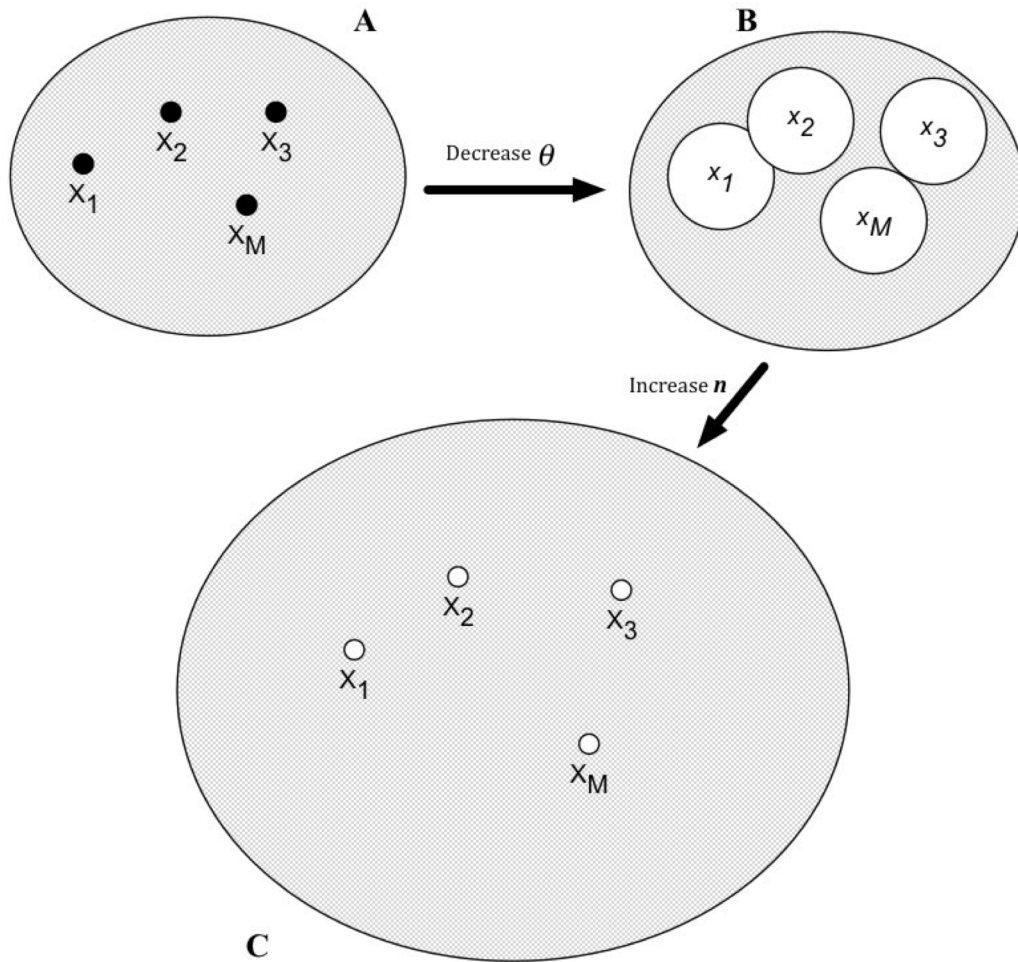
Figure 3. This figure illustrates the conceptual effect of decreasing θ, the dendritic threshold and increasing n, the overall population of cells. The large grey ovals denote the universe of possible patterns. The smaller circles within indicate M different dendritic segments, each representing one pattern within this universe. If θ is equal to the number of synapses s, as shown in oval A, very few patterns will match any of these segments (this is denoted by the small black circles). As you decrease θ the set of potential matches to each segment increases, as indicated by the large white circles in oval B. With smaller θ the segments will therefore be more tolerant to changes but there is a larger probability of false matches to random patterns. The ratio of white to grey becomes larger as you decrease θ. Oval C shows that when you increase n, the universe of possible patterns increases and the relative size of the white circles shrinks rapidly. Thus there is a tradeoff between these parameters. You get more robustness as you decrease θ at the cost of additional false positives. This is mitigated if you increase the overall cell population n.

In general a_t, the number of cells active in $\boldsymbol{A}_t$ will be greater than s, the number of synapses on a segment. s will thus represent a (possibly very small) subsample of the actual pattern. To get a sense of the actual numbers for realistic scenarios, consider the following example. Suppose $n = 10{,}000$ and $s = 30$. If the overall sparsity is 3% (i.e. $a_t = 300$ active cells) the segment is subsampling a tenth of the bits in A_t. With a threshold $\theta = 15$, the probability of a false match is about one in 10^{-15}. In other words, with 10% subsampling and with 50% noise robustness, there is a negligible chance of a false match with a different pattern. As we show later through simulations, the range of parameters for robust performance is fairly large. As long as n, s and θ are not trivially low numbers the recognition accuracy is extremely robust.

3.2.   Accuracy of a population of segments

Neurons have thousands of synapses arranged along numerous dendrite segments. It only takes 8 to 20 active synapses on a short section of dendrite to generate an NMDA spike (Major et al., 2013). Therefore a neuron has the potential to recognize hundreds of unique patterns. Further, in any cortical region, millions of neurons are each simultaneously trying to recognize hundreds of patterns. Will the robustness exhibited by a single dendrite segment be maintained in a region of neural tissue?

Consider the case of M independent dendritic segments (potentially on different cells), each with s synapses and a threshold of θ. In the strictest scenario, it is a false positive if any of the segments falsely detect a pattern.

Let S be a set of M vectors, $S = \{\boldsymbol{D}_0, \cdots, \boldsymbol{D}_{M-1}\}$ where each vector $\boldsymbol{D}_i$ represents a single dendritic segment. Given random presynaptic input $\boldsymbol{A}$, we classify it as belonging to this set as follows:

$$\boldsymbol{A} \in S \overset{\text{def}}{=} \exists_{\boldsymbol{D}_i \in S} match(\boldsymbol{D}_i, \boldsymbol{A}) = true \qquad (5)$$

Given a new pattern $\boldsymbol{A}_t$, how reliably can we classify it? Here we assume the number of noise bits to be $\leq s_i - \theta$ where $s_i = |\boldsymbol{D}_i|$. As such there are no false negatives, only false positives. The probability of a false positive (one or more matching segments) is:

$$P(\boldsymbol{A} \in S) = 1 - (1 - P(overlap(\boldsymbol{A}_t, \boldsymbol{D}_i) \geq \theta))^M \qquad (6)$$

In practice the probability of an individual overlap is extremely small and it is difficult to compute without numerical issues. It is useful to use instead the following bound:

$$P(\boldsymbol{A} \in S) \leq M \cdot P(overlap(\boldsymbol{A}_t, \boldsymbol{D}_i) \geq \theta) \qquad (7)$$

For sparse high dimensional vectors, with parameters in the ranges we are concerned with, Eq. (7) is in fact a very tight upper bound. To get a sense of the numbers and the overall capacity, consider the following example. Suppose n=10,000 and you have 3% sparsity ($a = 300$). By storing 30 synapses per segment using a segment threshold of 15, you can detect a million random SDR patterns with a false positive rate better than 1 in a billion!

This result points to a remarkable property of high dimensional sparse representations. You can convert patterns to a set of decorrelated high dimensional SDRs and simply store a small bit-wise subsample of each one. You can then classify a massive number of these patterns almost perfectly even in the presence of a large amount of noise and system unreliability. Thus a large collection of independent neurons, each with an independent set of segments, can robustly classify a very large number of patterns with a relatively small number of synapses.

3.3.   The surprising union property

We have shown that a small number of synapses can reliably detect patterns in large populations of cells. The synapses recognizing a given pattern have to be co-located on a dendritic segment. However it is highly unlikely that synapses on dendritic segments are cleanly segregated into individual patterns. It is much more likely that synapses are somewhat mixed together. This is consistent with experimental results which show that a dendritic segment can contain hundreds of

synapses yet a relatively small number of active synapses anywhere on the segment can trigger an NMDA spike (Major et al., 2013). In this section we consider what happens if the synapses to recognize multiple patterns are mixed together within a given dendritic segment. We show that our model neuron can maintain robust recognition when multiple sets of synapses dedicated to recognizing different patterns are mixed together on a common dendritic segment.

It turns out that one of the most fascinating properties of high dimensional sparse representations is their ability to reliably store a set of patterns within a single vector. We call this the "union property" as it involves creating a union (binary OR) of multiple patterns. Suppose we allow a single segment to contain s synapses from A_{t1} and an additional s synapses from A_{t2}. The segment would spike if at least θ cells from A_{t1} or A_{t2} are active. By adding additional groups of s synapses from other patterns, this segment will detect additional patterns. The vector representation of such a segment consists of the binary OR of the s synapses from each pattern. We say that this segment now represents a union of patterns and any combination of θ active synapses from this union will cause a spike.

The advantage of a union is that a fixed SDR element (such as that represented by a single segment) can be used to recognize a varying number of patterns. There will never be a false negative: the segment will reliably fire in the presence of any of the stored activity patterns with up to $s - \theta$ bits of noise. The downside of course is that there is now a larger potential for false positives. It is possible for the segment to spike due to a mixture of active cells, say half from A_{t1} and half from A_{t2}. As you add more patterns to the union there are an increasing number of mismatch possibilities and the segment is increasingly likely to spike for random patterns. Although forming a union introduces a potentially significant source of error, we will show that with high dimensional vectors pattern recognition can still be performed very reliably using such a union representation.

Formally, the mechanics of unions are simple. To "store" a set of M SDR vectors we simply take the Boolean OR of all of them to create a new binary vector $\mathbf{X}$ (Figure 4).

$$X = \bigcup_{i=0}^{M-1} x_i \tag{8}$$

Some of the bits in x_i may overlap so X is now a binary vector such that $|X| \leq \sum_{i=0}^{M-1} |x_i|$. To check if a new pattern $\mathbf{y}$ is a member of the set, we compute the match as in Eq. (1).

How reliable is the classification operation? We first consider the case of exact matches, i.e. $\theta = |y|$. For simplicity we also assume all vectors in X have the same number of ON bits. Note that if $y = x_i$ for some i, the match operation will always be successful. However for other vectors there is a chance of a false positive match due to mix and match errors. A false positive with a new random pattern y occurs if all of the bits in y overlap with X. When $M = 1$, the probability that any given bit is 0 is $1 - q$, where $q = \frac{|x_0|}{n}$. As M grows, the probability that a given bit is still 0 is:

$$p_0 = (1 - q)^M \tag{9}$$

The probability that a given bit in $\mathbf{X}$ is ON is therefore $(1 - p_o)$. The probability of a false positive match, i.e. that all the bits in $\mathbf{y}$ are ON is therefore:

$$P(overlap(X, y) \geq \theta) = (1 - p_0)^{|y|} = (1 - (1 - q)^M)^{|y|} \tag{10}$$

This calculation is equivalent to the probability of false positives in Bloom filters (Bloom, 1970). We can now plug this back into Eq. (4) to compute the probability of error for inexact matches. After M union operations, the expected number of ON bits in X is:

$$E[|\mathbf{X}|] = (1 - p_0)n \tag{11}$$

The expected size of the overlap set is:

$$E[|\Omega_X(n, a, b)|] = \binom{(1 - p_0)n}{b} \times \binom{n - (1 - p_0)n}{a - b} \tag{12}$$

We can now calculate how accurately a single dendritic segment can represent a mix of different patterns. Suppose a segment **D** represents a union of M different patterns. The expected number of synapses on this segment s is $E[s] = (1 - p_0)n$. Given random presynaptic activity $\mathbf{A}_t$, the expected probability of a false match is therefore:

$$E[P(overlap(\mathbf{D}, \mathbf{A}_t) \geq \theta)] = \frac{\sum_{b=\theta}^{s} E[|\Omega_D(n, a, b)|]}{\binom{n}{a}} \tag{13}$$

The equation is complex but the numbers can be illuminating. As an example, suppose the population size $n = 20000$, with $a = 100$ cells active at a time. Suppose each pattern on a segment is represented by 25 synapses and $\theta = 15$. If you union together $M = 10$ patterns, on average you would get fewer than 250 synapses on the segment. In this scenario the false positive rate is less than one in 10^{11}. To gain an intuition for this, it is useful to think about the expected number of ON bits in the SDR, i.e. Eq. (11). With $M = 10$, 98.75% of the bits in the vector representing this segment are zero. When you match against an additional vector, there is a very high chance it will have most of its bits among this 98.75%, and hence it won't lead to a false positive. Only vectors that have at least 15 of their bits ON among the 1.5% will cause a false positive.

The net impact of the union property is that individual dendritic segments can be sloppy and reliable at the same time. There are limits, but such segments can mix together a number of independent patterns with virtually no chance of false positive errors.

4. Simulation Results

Due to the various factorials and exponentials involved in the above equations it is sometimes difficult to develop a solid intuitive understanding of the various scaling properties. The range of parameters leading to robust recognition may be unclear, and it may also be unclear whether these results translate to real neurons and experimental data. In this section we describe a number of simulation results. Our goal is to develop intuitions for the reader, and to demonstrate the applicability of the results.

We ran a number of simulations to illustrate some of the properties of dendritic matching (i.e. Eq. (4)). Figure 4A shows the effect of the underlying dimensionality of the representation space, i.e. the population of cells. We plot the drop in error rates as you increase the population size, n, while maintaining a fixed sparsity level and a fixed number of synapses on a segment. The error drops rapidly (faster than exponentially) as n increases, becoming essentially 0 once $n > 2000$. Note that it is not possible to get robust recognition with a dense representation, as shown by the dashed line representing a 50% activity level. Thus both sparsity and high dimensionality are required to achieve robust recognition with a small number of synapses.
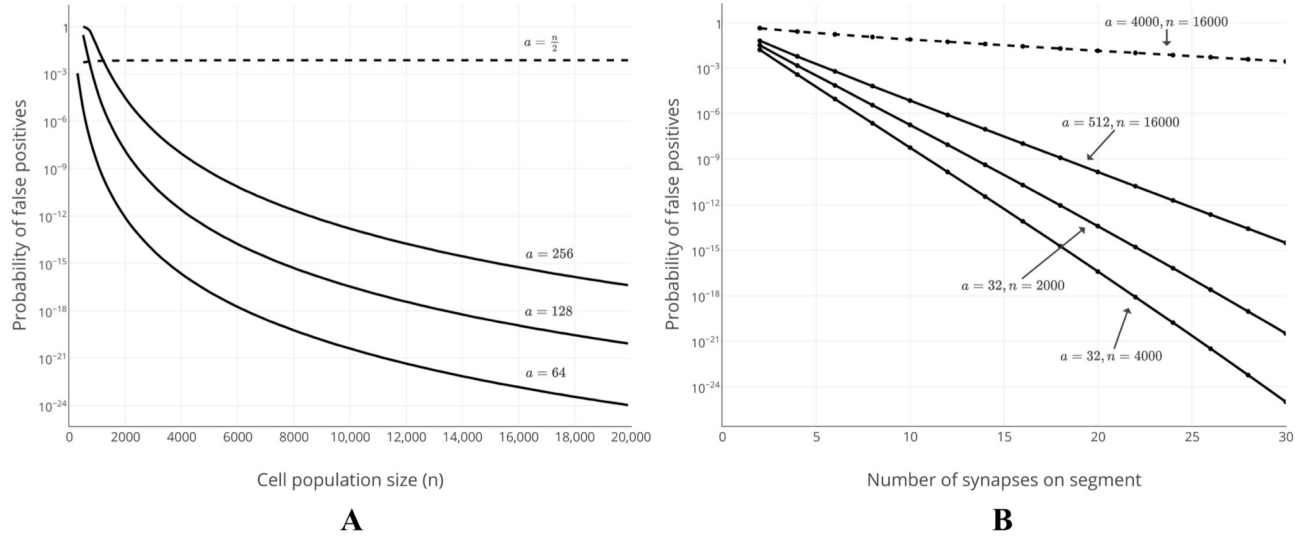
Figure 4. **A.** This graph illustrates the behavior of Eq. (4) and the effects of cell population and sparsity. The three solid curves show the rapid drop in error rates as the number of cells n increases. Each curve shows a different sparsity level. For example, if 128 out of 4000 cells are active (3.2% sparsity) the error rate is a little higher than 10^{-12}. The dashed line corresponds to an activity level of 50%. The fact that the error corresponding to this condition does not drop demonstrates that both sparsity and high dimensionality are required to achieve low error rates. In all of these simulations, the number of synapses $s=24$ and the dendritic threshold $\theta=12$, corresponding to 50% noise tolerance. **B.** This graph illustrates the behavior of Eq. (4) and the effect of s, the number of synapses on a segment. The segment threshold is set to $\theta = \frac{s}{2}$. As the curves show, there is an exponential improvement in error rates as the number of synapses increases. The three solid curves show error rates for a variety of sparsity levels ($\frac{a}{n}$ is between 0.8% to 3.2%) and a range of dimensionalities (n between 2000 and 16000). The curves show that if the overall activity level is sparse and the cell population is sufficiently high, values of s between 15 and 25 can lead to low error rates. This helps explain why even a tiny number of synapses on a dendritic segment subsampling from a much larger pattern of activity is sufficient for robust recognition performance. The dashed curve is an example illustrating that a relatively dense representation (25% shown) does not lead to low error even if the underlying dimensionality is high.

Figure 4B examines the effect of changing the number of synapses on a dendritic segment. It illustrates how segments can store a tiny subsample from a large population and still robustly recognize complex patterns. If the population of cells increases beyond a few thousand, and the overall activity is sparse, it is possible to achieve reliable recognition with a small sample of each pattern. The chart shows that the error rate decreases exponentially with the number of synapses. For many situations a subsample of 20 to 25 synapses leads to an error rate better than 10^{-10}. Note that since $\theta = \frac{s}{2}$, this includes a noise level up to 50%. This helps explain how even a small number of synapses on a segment are sufficient for robust recognition performance. The dashed line shows that denser representations lead to high error rates. Although not shown, small a and n (e.g. $a = 32$ and $n = 128$) also lead to high error. In other words, in order to achieve accurate recognition with a small subsample, you need both sparsity and a sufficiently high dimensionality.

The mathematics behind sparse representations can help provide insight into key experimental results. In Figure 5 we show that the equations can be used to explore the effect of different dendritic spike thresholds and suggest an optimal range. Figure 5 shows the median probability of error as a function of the synaptic threshold for a dendritic spike. Each point on the graph holds the threshold θ fixed and represents the median probability of error computed over a large range of all other parameters: n (the number of potential synapses), presynaptic activity a, and s the number of synapses on each segment. We systematically varied n from 10,000 to 200,000, presynaptic activity
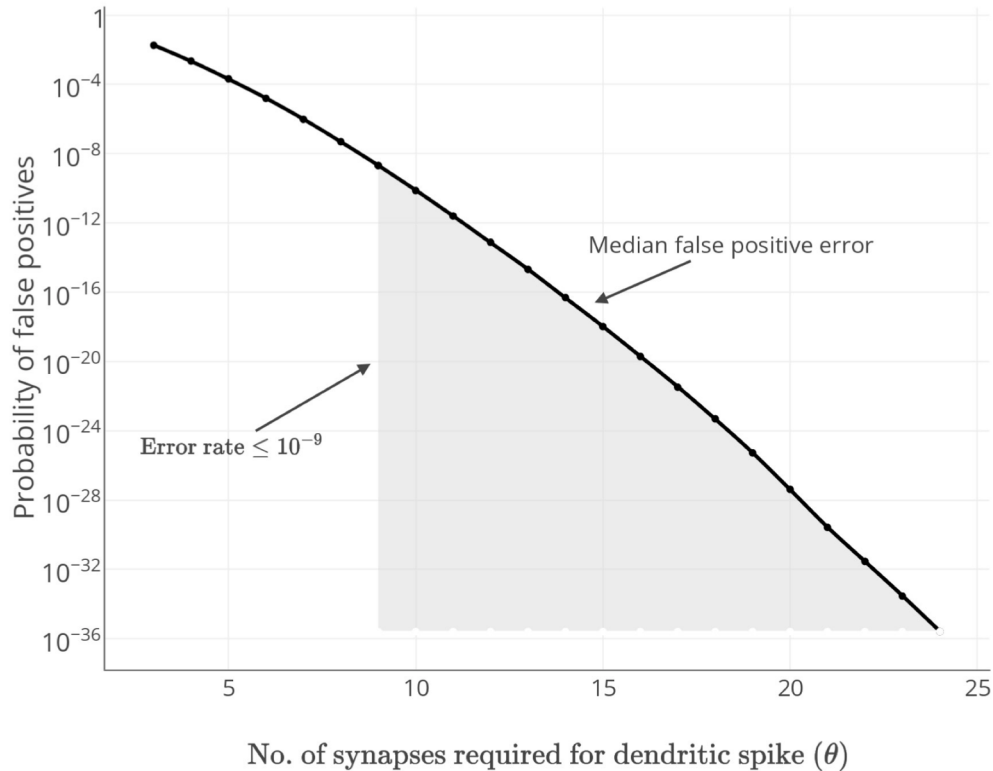
Figure 5. The equations can be used to characterize the effect of different synaptic thresholds and predict an ideal range. This chart shows the median probability of error as a function of the synaptic threshold for a dendritic spike θ. Each point on the graph holds θ fixed and represents the median probability of error computed over a large range of all other parameters: n (the number of potential synapses), presynaptic activity a, and s, the number of synapses on each segment. We exhaustively varied n from 10,000 to 200,000, presynaptic activity from 0.5% to 3% of n, and s from 20 to 50. The shaded area of the chart is the region corresponding to low error (for illustration purposes we show the range of thresholds that lead to an error of 1 in a billion or lower). The chart demonstrates that a spike threshold of 9 and higher leads to low error rates under a very wide range of assumptions. A spike threshold beyond 15 or 20 leads to diminishing returns. Beyond that the error rates are so low that the extra metabolic cost of additional synapses due to higher thresholds is not justified.

from 0.5% to 3% of n, and s from 20 to 50. The shaded area of the chart is the region corresponding to low error (for illustration purposes we show the range of thresholds that lead to an error of 1 in a billion or lower).

Figure 5 demonstrates that a NMDA spike threshold of 9 and higher leads to low error rates under a very wide range of assumptions. A spike threshold beyond 15 or 20 leads to diminishing returns. Beyond 20 the error rates are so low that the extra metabolic cost of forming additional synapses due to higher thresholds is not justified. Thus the equations predict that the optimal dendritic spiking threshold is between 9 and 20. This lines up well with experimental results, which show actual spiking thresholds to be between 8 and 20 (Major et al., 2013; Branco and Häusser, 2011).

Figure 6 shows two simulations that demonstrate the effect of the union property on a dendritic segment. As discussed earlier, a segment can contain synapses from a mixture of independent patterns and initiate an NMDA spike if any of them are detected. Figure 6A shows how the expected number of synapses on a segment scales with the number of patterns (i.e. Eq. (11)). Experimental results show that a single dendritic segment can contain anywhere from 100 and 400 synapses (Major et al., 2013). The graph suggests that this translates to between 4 and 16 independent patterns, dependent on the number of synapses used to represent each pattern. Figure 6B illustrates the
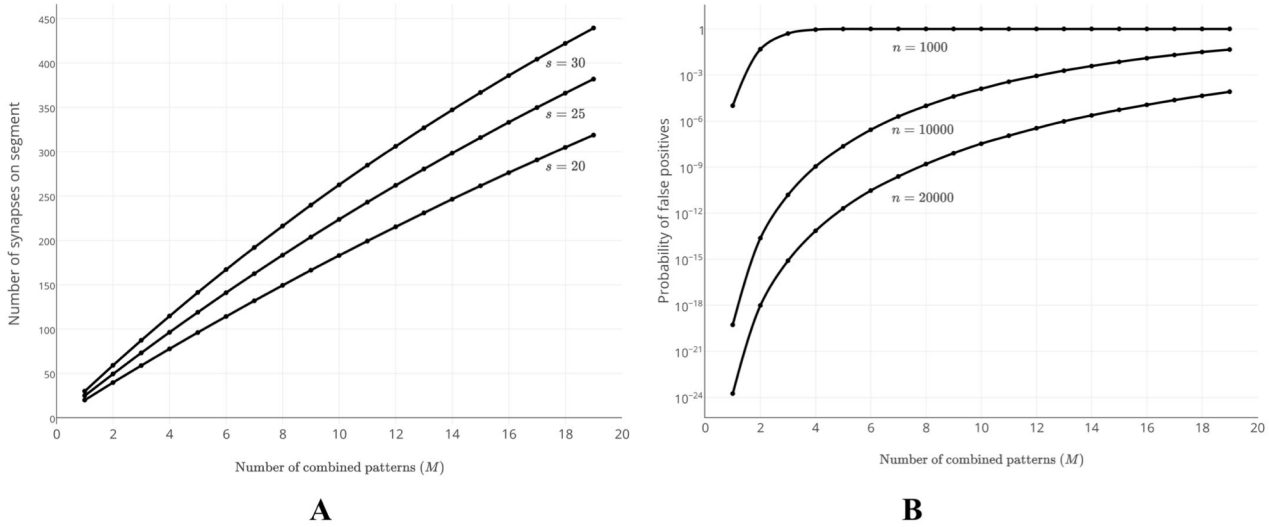
Figure 6. **A**. This figure plots the expected number of synapses on a segment that represents M different patterns. Each curve shows a different value of *s*, the number of synapses used represent a single pattern. The point of this plot is to show the curve grows slower than linearly because as more patterns are included in the union the chance that two patterns contain overlapping presynaptic cells grows. **B**. This figure shows the expected error rates for segments that contain a union of multiple patterns. The error increases monotonically with the number of patterns "stored" on each segment. The size of the presynaptic population is a critical factor; a larger presynaptic population implies a much lower probability for mix and match errors. A small presynaptic population of 1000 leads to relatively high error rates, but a population of 20,000 can lead to extremely low error rates even with 10 patterns stored. In this graph the activity of the population, *a*, is *200*. *s*, the number of synapses representing each pattern, is *25*. The dendritic threshold θ is *15*.

probability of mix and match errors for segments that represent such a union of patterns. Here n, the size of the presynaptic population, and the overall sparsity levels are critical factors. As the graph shows, a larger presynaptic population implies a low chance of mix and match errors. A small presynaptic population of 1000 leads to relatively high error rates, but a population of 20,000 can lead to extremely low error rates even with 10 patterns stored. Overall these graphs show that it is possible for a single dendritic segment to sloppily store a mixture of multiple patterns, yet maintain remarkably high accuracy for detecting each pattern.

Note that the error rates plotted in Figures 4, 5, and 6B are for individual neurons. However, the correct classification of a pattern is always performed by a population of neurons. Populations of active cells can contain a significant number of incorrect false positives activations without error in classification of the entire population. Therefore, even if individual neurons operate in a region of unacceptably high false positives, the population accuracy will be substantially lower.

5.   Discussion

Sparse distributed representations are ubiquitous in neocortex. In this paper we have proposed a formal mathematical model for sparse representations in neocortex based on a specific neuron model and associated operations. Our model neuron and the core operations of overlap and match are inspired by experimental findings on active dendritic processing and NMDA spikes in pyramidal neurons. We derived a number of scaling laws demonstrating that systems based on these principles can achieve extreme robustness to noise and faults in the system. Our simulation results provide insights into various parameter regimes and show that both sparsity and high dimensionality are required for maximum accuracies. In addition we show that the equations can be used to predict experimental results, such as the optimal spiking thresholds for active dendrites.

Our work is related to the theoretical work of (Kanerva, 1988; Olshausen and Field, 2004) as well as the work on active dendrites (Poirazi and Mel, 2001). A major difference with these papers is that they use a typical Euclidean distance norm and a weighted linear sum with scalar vectors instead of our overlap metric with binary vectors. (Babadi and Sompolinsky, 2014) have also developed a theoretical model of sparsity. Although they too conclude that sparsity in high dimensional spaces is desirable, their neuron model is simplistic and does not consider the realities of active dendrites nor does it use overlap. The overlap metric matches the biological operation in neurons and is key to the robustness properties described above. Overall the numerical results in our paper more closely match known biological data. We also explicitly characterize error rates and show how to achieve high reliability under very noisy and imprecise scenarios. Despite the above differences, taken together these papers provide general support for the power of sparse high dimensional representations, and the insights to be gained from mathematical modeling and simulations of sparsity.

In this paper we have discussed neurons with active dendrites, and their error rates in pattern detection. We have not discussed the functional implication of these matches. What does the neuron do once a match occurs? The functional benefits of active dendrites are a topic of active research, and there have been a few theories proposed in the literature. These theories include the efficient propagation of neural activity (Polsky et al., 2009), facilitating top-down prediction (Larkum, 2013), sequence storage (Losonczy et al., 2008), and gain control (Larkum et al., 2004). Elsewhere we have described a detailed theory and working implementation that shows how active dendrites and networks of pyramidal cells lead to a sophisticated and practical sequence memory algorithm (Hawkins and Ahmad, 2015). The mathematical framework proposed in this paper can be used to provide insights into the overall robustness and capabilities of these models.

The equations in this paper assume random and de-correlated neural activity. The distribution of individual spiking neurons and neural correlation in neocortex is a topic of some debate (Cohen and Kohn, 2011). A number of papers have suggested that one of the outcomes of neural plasticity and inhibition is to de-correlate the inputs. For example, Hebbian-style learning plus inhibition leads to individual neurons that represent successive principal components of the input space (Oja, 1982). Indeed in-vivo measurements of neural activity imply correlation is low, even for neurons with highly overlapping receptive fields (Smith and Häusser, 2010; Ecker et al., 2010). It is our belief that it is not necessary to have completely uniform random distributions to obtain high fault tolerance. Extending our results to other distributions is an interesting topic for future work.

The equations in this paper assume binary synapses and binary cell activations, and ignore the possibility of scalar weights or outputs. The question of binary synapses is a topic that is heavily debated in literature. For example some researchers have argued that synapses are inherently binary (Petersen et al., 1998), while others have argued the opposite (Enoki et al., 2009). Interestingly, in their model (Poirazi and Mel, 2001) found that binary receptive fields led to the best memory capacity. We simply note that the robustness demonstrated in this paper does not rely on continuous values. Continuous vectors are a superset of binary vectors and thus in theory can only add to the power demonstrated here. A study of the scaling properties of continuous SDRs is beyond the scope of this paper, but represents an interesting direction for future research.

Individual neurons and synapses are inherently unreliable (Faisal et al., 2008), yet the overall system works extremely well. The results in this paper shed light on how cortical processing can be incredibly robust and fault tolerant as long as the underlying representation is sparse and high dimensional. The mathematical model described here is not specific to any sensory modality or cortical area. As such the properties should be ubiquitous for pyramidal cells everywhere. It is our

hope that over time, a complete theoretical understanding of the learning and scaling properties of neocortical representations can be developed.

Acknowledgments

To be written.

References

Antic, S. D., Zhou, W. L., Moore, A. R., Short, S. M., and Ikonomu, K. D. (2010). The decade of the dendritic NMDA spike. *J. Neurosci. Res.* 88, 2991–3001. doi:10.1002/jnr.22444.

Babadi, B., and Sompolinsky, H. (2014). Sparseness and Expansion in Sensory Representations. *Neuron* 83, 1213–1226. doi:10.1016/j.neuron.2014.07.035.

Barth, A. L., and Poulet, J. F. a (2012). Experimental evidence for sparse firing in the neocortex. *Trends Neurosci.* 35, 345–355. doi:10.1016/j.tins.2012.03.008.

Bloom, B. H. (1970). Space/time trade-offs in hash coding with allowable errors. *Commun. ACM* 13, 422–426. doi:10.1145/362686.362692.

Branco, T., and Häusser, M. (2011). Synaptic integration gradients in single cortical pyramidal cell dendrites. *Neuron* 69, 885–92. doi:10.1016/j.neuron.2011.02.006.

Chklovskii, D. B., Mel, B. W., and Svoboda, K. (2004). Cortical rewiring and information storage. *Nature* 431, 782–8. doi:10.1038/nature03012.

Cohen, M. R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nat. Neurosci.* 14, 811–819. doi:10.1038/nn.2842.

Crochet, S., Poulet, J. F. a, Kremer, Y., and Petersen, C. C. H. (2011). Synaptic mechanisms underlying sparse coding of active touch. *Neuron* 69, 1160–1175. doi:10.1016/j.neuron.2011.02.022.

Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., Logothetis, N. K., and Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science* 327, 584–587. doi:10.1126/science.1179867.

Enoki, R., Hu, Y. ling, Hamilton, D., and Fine, A. (2009). Expression of Long-Term Plasticity at Individual Synapses in Hippocampus Is Graded, Bidirectional, and Mainly Presynaptic: Optical Quantal Analysis. *Neuron* 62, 242–253. doi:10.1016/j.neuron.2009.02.026.

Faisal, A. A., Selen, L. P. J., and Wolpert, D. M. (2008). Noise in the nervous system. *Nat. Rev. Neurosci.* 9, 292–303. doi:10.1038/nrn2258.

Gazzaniga, M. S. [Ed] (2004). *The cognitive neurosciences (3rd ed.).* doi:10.1136/bmj.312.7024.193.

Graziano, M. S. A., and Aflalo, T. N. (2007). Mapping behavioral repertoire onto the cortex. *Neuron* 56, 239–251. doi:10.1016/j.neuron.2007.09.013.

Graziano, M. S. A., Taylor, C. S. R., and Moore, T. (2002). Complex movements evoked by microstimulation of precentral cortex. *Neuron* 34, 841–851. doi:10.1016/S0896-6273(02)00698-0.

Hawkins, J., and Ahmad, S. (2015). Why Neurons Have Thousands of Synapses, A Theory of Sequence Memory in Neocortex. Available at: http://arxiv.org/abs/1511.00083.

Hromádka, T., DeWeese, M. R., and Zador, A. M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol.* 6, 0124–0137. doi:10.1371/journal.pbio.0060016.

Kanerva, P. (1988). *Sparse Distributed Memory*. The MIT Press.

Kiani, R., Esteky, H., Mirpour, K., and Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J. Neurophysiol.* 97, 4296–4309. doi:10.1152/jn.00024.2007.

Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.* 36, 141–51. doi:10.1016/j.tins.2012.11.006.

Larkum, M. E., Nevian, T., Sandler, M., Polsky, A., and Schiller, J. (2009). Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. *Science* 325, 756–760. doi:10.1126/science.1171958.

Larkum, M. E., Senn, W., and Lüscher, H. R. (2004). Top-down dendritic input increases the gain of layer 5 pyramidal neurons. *Cereb. Cortex* 14, 1059–1070. doi:10.1093/cercor/bhh065.

Legenstein, R., and Maass, W. (2011). Branch-specific plasticity enables self-organization of nonlinear computation in single neurons. *J. Neurosci.* 31, 10787–802. doi:10.1523/JNEUROSCI.5684-10.2011.

Losonczy, A., Makara, J. K., and Magee, J. C. (2008). Compartmentalized dendritic plasticity and input feature storage in neurons. *Nature* 452, 436–41. doi:10.1038/nature06725.

Major, G., Larkum, M. E., and Schiller, J. (2013). Active properties of neocortical pyramidal neuron dendrites. *Annu. Rev. Neurosci.* 36, 1–24. doi:10.1146/annurev-neuro-062111-150343.

McCulloch, W. S., and Pitts, W. H. (1943). A logical calculus of ideas imminent in nervous activity. *Bull. Math. Biophys.* 5, 115–133. doi:10.1007/BF02478259.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *J. Math. Biol.* 15, 267–273. doi:10.1007/BF00275687.

Olshausen, B. A., and Field, D. J. (2004). Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* 14, 481–487. doi:10.1016/j.conb.2004.07.007.

Olshausen, B. A., and Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res.* 37, 3311–3325. doi:10.1016/S0042-6989(97)00169-7.

Petersen, C. C., Malenka, R. C., Nicoll, R. A., and Hopfield, J. J. (1998). All-or-none potentiation at CA3-CA1 synapses. *Proc. Natl. Acad. Sci. U. S. A.* 95, 4732–4737. doi:10.1073/pnas.95.8.4732.

Poirazi, P., and Mel, B. W. (2000). Choice and Value Flexibility Jointly Contribute to the Capacity of a Subsampled Quadratic Classifier. *Neural Comput.* 12, 1189–1205. doi:10.1162/089976600300015556.

Poirazi, P., and Mel, B. W. (2001). Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. *Neuron* 29, 779–796. doi:10.1016/S0896-6273(01)00252-5.

Polsky, A., Mel, B., and Schiller, J. (2009). Encoding and decoding bursts by NMDA spikes in basal dendrites of layer 5 pyramidal neurons. *J. Neurosci.* 29, 11891–11903. doi:10.1523/JNEUROSCI.5250-08.2009.

Smith, S. L., and Häusser, M. (2010). Parallel processing of visual space by neighboring neurons in mouse visual cortex. *Nat. Neurosci.* 13, 1144–1149. doi:10.1038/nn.2620.

Spruston, N. (2008). Pyramidal neurons: dendritic structure and synaptic integration. *Nat. Rev. Neurosci.* 9, 206–221. doi:10.1038/nrn2286.

Vinje, W. E., and Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science (80-.).* 287, 1273–6. doi:10.1126/science.287.5456.1273.

Weliky, M., Fiser, J., Hunt, R. H., and Wagner, D. N. (2003). Coding of natural scenes in primary visual cortex. *Neuron* 37, 703–718. doi:10.1016/S0896-6273(03)00022-9.

Wu, X. E., and Mel, B. W. (2009). Capacity-enhancing synaptic learning rules in a medial temporal lobe online learning model. *Neuron* 62, 31–41. doi:10.1016/j.neuron.2009.02.021.