Perceptron learning and classification in a modeled cortical pyramidal cell

Toviah Moldwin[1], Idan Segev[1,2]

[1]Department of Neurobiology and [2]Edmond and Lily Safra Center for Brain Sciences, the Hebrew University of Jerusalem, Jerusalem, Israel

Correspondence:
Toviah Moldwin
Toviah.moldwin@mail.huji.ac.il

Keywords: Compartmental modeling, Nonlinear dendrites, Cortical excitatory synapses, Single neuron computation, Machine learning, Synaptic weights, Dendritic voltage attenuation

Abstract

The perceptron learning algorithm and its multiple-layer extension, the backpropagation algorithm, are the foundations of the present-day machine learning revolution. However, these algorithms utilize a highly simplified mathematical abstraction of a neuron; it is not clear to what extent real biophysical neurons with morphologically-extended nonlinear dendritic trees and conductance-based synapses could realize perceptron-like learning. Here we implemented the perceptron learning algorithm in a realistic biophysical model of a layer 5 cortical pyramidal cell. We tested this biophysical perceptron (BP) on a memorization task, where it needs to correctly binarily classify 1000 patterns, and a generalization task, where it should discriminate between "noisy" patterns drawn from one of two probability distributions. We show that, for most cases, the BP performs these tasks with an accuracy comparable to that of the original perceptron. We concluded that cortical pyramidal neurons can act as powerful classification devices.

Introduction

There has been a long-standing debate within the neuroscience community about the existence of 'grandmother neurons', or individual cells that code for high-level concepts, such as a person's grandmother. Recent experimental evidence, however, has indicated that there indeed seem to be units that are selective to specific high-level inputs. In particular (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005) found cells in the human medial temporal lobe (MTL) that fire in response to images of a particular celebrity, such as Jennifer Aniston or Halle Berry. One remarkable aspect of this finding is that different images of the same celebrity would elicit a response in these neurons even if the subject of the image was facing a different direction, wearing different clothes, or under different lighting conditions. In other words, the specificity of these MTL cells is invariant to certain transformations of the sensory stimulus. Regardless of whether this finding is evidence for grandmother cells or merely for sparse coding (Quiroga, Kreiman, Koch, & Fried, 2008), it is apparent that individual neurons can be highly selective for a particular pattern of sensory input and also possess a certain level of generalization ability or "tolerance" to differences in the input that do not change the essence of the sensory scene.

From a physiological standpoint, achieving a high degree of accuracy on a recognition task is a daunting challenge for a single neuron. To put this in concrete terms, a pyramidal neuron in rat hippocampal CA1 receives around 30,000 excitatory synapses and 1,700 inhibitory synapses (Megías, Emri, Freund, & Gulyás, 2001). As a first approximation, at any given time, each of this neuron's presynaptic inputs can either be active or inactive, yielding $2^{31,700}$ possible binary patterns. If the presynaptic inputs contain information about low-level sensory stimuli (such as pixels or orientation filters) and the postsynaptic neuron needs to respond only to images of Jennifer Aniston, there must be some physiological decision procedure by which the neuron "chooses" which of those $2^{31,700}$ patterns are sufficiently close to the binary representation of Jennifer Aniston to warrant firing an action potential. There are several ways that a neuron can selectively respond to different input patterns. The most well-known method is to adjust synaptic "weights" such that only input patterns which activate a large number of highly-weighted synapses will cause the cell to fire. It is this principle which serves as the basis of the perceptron learning rule (Rosenblatt, 1958) which is, in turn, the foundation for the artificial neural networks (ANNs) that are commonly used today in machine learning and deep networks (Krizhevsky, Sutskever, & Hinton, 2012; Rumelhart, Hinton, & Williams, 1986).

The perceptron utilizes a mathematical abstraction of a neuron which applies a nonlinear function (such as a sigmoid) to the weighted sum of its input (Figure 1A). This abstraction is known as the McCulloch and Pitts (M&P) neuron (McCulloch & Pitts, 1943). The nonlinear output of the neuron plays the role of a classifier by producing a positive output (a spike, +1) in response to some input patterns and a negative output (no spike, -1) in response to other patterns. This "perceptron" neuron model is trained in a supervised manner by being given training examples which are labeled as belonging to either the positive or negative category. The perceptron output is calculated for each example, and if the perceptron output for a particular example does not match the label, the perceptron's weights are updated such that its output will be closer to the correct output for that example in the future.

While the remarkable efficacy of networks of M&P neurons has demonstrated for various learning tasks, few attempts have been made to replicate the perceptron learning algorithm in a detailed biophysical neuron model with a full morphology and active dendrites with conductance-based synapses. Some work (Safaryan, Maex, Davey, Adams, & Steuber, 2017; Steuber et al., 2007) has been done in this regard for Purkinje cells, which have been classically conceived of as perceptrons (Albus, 1971; Marr, 1969). These studies demonstrated that detailed models of Purkinje cells can learn to discriminate between different patterns of input from the parallel fibers (PF) via a perceptron-like usage of long-term depression (LTD), which is known to occur in PF-Purkinje synapses. Crucially, the difference between the Purkinje cell's responses to learned versus unlearned patterns was the duration of the pause between spikes in the Purkinje cell's output subsequent to the presentation of PF input. Steuber et al. (Steuber et al., 2007) argue that this pause duration-based learning depends on mediation of calcium concentrations inside the cell. This is different from the more direct M&P-like mechanism of synapses being weighted such that only certain input patterns will reach the cell's spiking threshold.

It thus remains to be determined whether other cells in the brain - which may not have this calcium-dependent pause-duration learning - can integrate and classify their inputs in a perceptron-like manner. This is a question of particular interest for pyramidal neurons, which comprise most excitatory neurons in the cortex. A variety of nonlinear models exist to describe learning in pyramidal cells, such as the Clusteron (Mel, 1991, 1992) and the two-layer model (Poirazi, Brannon, & Mel, 2003). These msodels take advantage of the superlinearities present in pyramidal cells dendrites, especially the NMDA-dependent synergistic coactivation of nearby excitatory synapses. However, the existence of nonlinearities does not necessarily preclude the possibility of perceptron-like linear learning algorithm for pyramidal cells, especially given that a linear model seems to capture a large percent ($r^2 = 0.82$) of the variance in output of a pyramidal cell (Poirazi et al., 2003). Still, given that pyramidal neurons synapses implement a Hebbian-like

learning rule of spike timing-dependent plasticity (STDP) (Markram, Lübke, Frotscher, & Sakmann, 1997; Zilberter et al., 2009), the perceptron learning rule remains an attractive candidate for learning and plasticity in real cortical neurons. Moreover, although the two-layer model of Poirazi, Brannon, & Mel, 2003 describes how pyramidal cells might learn and compute functions of their inputs, there has not, as of yet, been a successful demonstration that learning can be implemented in a two-layer manner in a detailed biophysical model of a pyramidal cell. (Mel, 1992) did show an example of learning via clustering in a detailed biophysical model, but that learning paradigm did not make use of modifying synaptic weights and thus only made use of second-order information about input correlations and mostly disregarded first-order information about the significance of the individual inputs themselves.

In this study, we used the perceptron learning algorithm to teach a detailed realistic biophysical model of a layer 5 pyramidal cell (Hay et al., 2011) to solve two kinds of classification problems: a memorization task, where the neuron must correctly classify a predefined set of input patterns, and a generalization task, in which the neuron has to determine which of two probability distributions an input came from (e.g., men vs. women). We explored the effect of real neurons with extended nonlinear dendritic tree and conductance-based excitatory synapses on their ability to perform classification tasks of the sort commonly solved by artificial neurons (See **Discussion** for a treatment of why only excitatory synapses were used). We found that the performance of the biophysical perceptron (BP) is close to that of the artificial counterpart.

Results

Memorization task

To implement the perceptron learning algorithm with a detailed biophysical model of an layer 5 thick tufted pyramidal cell (L5PC) we distributed conductance-based excitatory (AMPA/NMDA) synapses on the detailed model by (Hay, Hill, Schürmann, Markram, & Segev, 2011). We varied the total number of synapses (N) placed on the modelled neuron in order to determine its classification capacity. We tested conditions of N = 500, N = 1000, and N = 10,000 (we varied the number of synapses, as opposed to the number of patterns, due to computational constraints). Binary patterns were created by randomly selecting 200 out of the N excitatory synapses to be activated in a given pattern (Figure 1B). For each value of N, 1000 binary patterns were created and evenly divided into a “positive” (+1) group (for which the neuron should produce at least one spike) and a “negative” (-1) group (for which the neuron should not produce a spike). To achieve perfect accuracy, the neuron would have to fire in response to 500 out of the 1000 patterns and not fire in response to the other 500 patterns. Note that initially, there is no reason for the neuron to perform at better than chance level, because all the patterns contain the same number of active synapses.

We then used the perceptron learning algorithm (see **Methods**) to modify the synaptic weights such that the cell could correctly classify all the patterns (Figure 1C). This procedure was repeated in conditions in which synapses were placed over the whole dendritic tree, only on the apical tuft, only on the basal tree, or only on the soma to determine how the location of the synapses would affect the cell’s ability to classify patterns using the perceptron learning rule. For each of these conditions of synaptic distribution, we ran the algorithm for the same cell with passive dendrites (i.e. no voltage-dependent NMDA synapses or voltage-dependent dendritic ion channels) to determine how the active properties of the cell might affect its performance as a perceptron. We also tested the algorithm with current-based synapses rather than of conductance-based synapses, to examine whether conductance synapses has any advantages or disadvantages with respect to the cell’s performance as a classifier.

Figure 2 shows the learning curves (Figure 2A) and classification accuracy (Figure 2B) for each of the above-mentioned conditions. In all cases the cell is able to improve its performance relative to chance, indicating that the complexity of biophysical cells does not preclude perceptron

learning despite the fact that the learning algorithm having been devised for a much simpler abstraction of a cell.

We compared the classification accuracy for each condition in the biophysical model to an equivalent M&P perceptron (see **Methods**). When all synapses are placed on the soma or the proximal basal tree of the biophysical perceptron, the classification accuracy of the biophysical perceptron is near to that of the M&P perceptron. Removing NMDA conductances and active dendritic ion channels did not substantively affect this result.

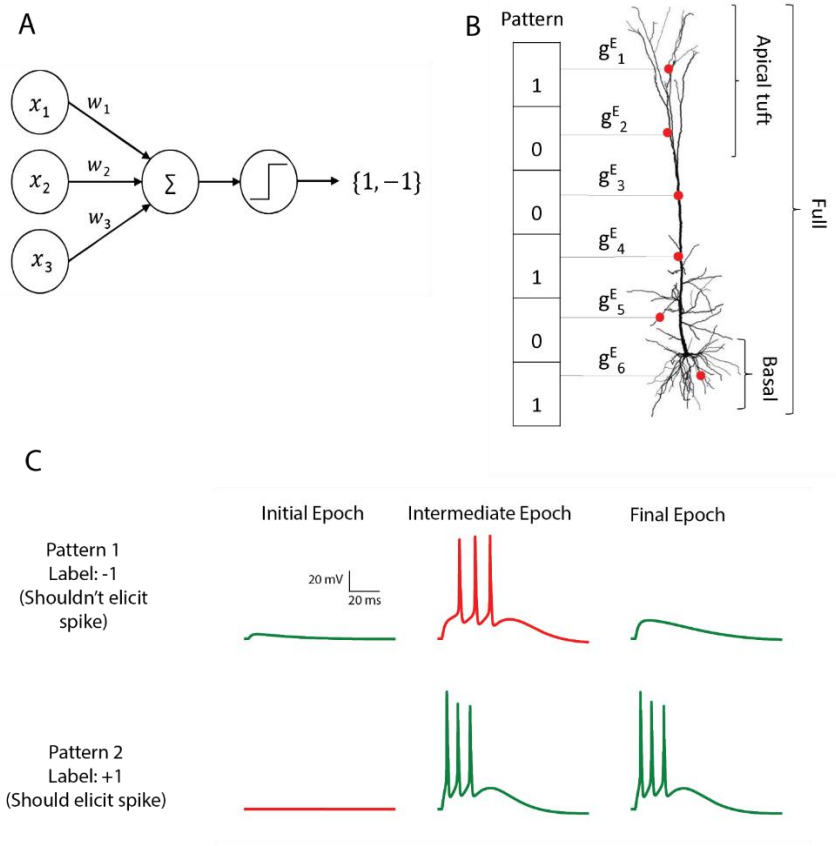


Figure 1. The M&P and biophysical perceptron
(A) The M&P perceptron. The presynaptic cells are represented by their firing rates, $x_1, x_2, x_3, \ldots x_i$, each of which is multiplied by the respective synaptic weight $w_1, w_2, w_3, \ldots w_i$ and then summed together with the other inputs. The perceptron produces an output of +1 if the weighted sum of all inputs is greater than a threshold and -1 otherwise. The task of the perceptron is to learn the appropriate synaptic weights, such that it will produce an output of +1 for an arbitrarily predefine subset of x_i, and -1 for the remaining subset of x_i.
(B) Schematic of the biophysical perceptron. A layer 5 pyramidal cell model with excitatory synapses (red dots) receiving presynaptic input patterns. The synaptic weights are the excitatory conductance, $g^E{}_i$, for the respective synapse, i. In this model, a presynaptic input pattern consists of a particular set of synaptic inputs that are either active "1" or inactive "0.
(C) An example of the learning process in the biophysical perceptron. Two input patterns, each with 1000 synapses, are presented to the model neuron. for pattern 1 the model cell should not generate any spike, whereas for pattern 2 it should. In the initial epoch neither pattern elicits a spike (traces at left). The output for the pattern 1 is thus correct (green trace) but incorrect (red trace) for pattern 2. In an intermediate epoch of the learning algorithm (middle column), the neuron has increased its synaptic conductances sufficiently so that pattern 2 does elicit spikes, however pattern 1 also produces spikes (but it should not). By the final epoch (left column), the weights are adjusted such that the neuron correctly classifies the two patterns (right column).

As expected from the theoretical literature (Chapeton, Fares, LaSota, & Stepanyants, 2012), the accuracy in each condition increases with the number of synapses comprising each pattern. This can be seen in Figure 2B, where the classification accuracy can be seen to improve in each condition as we move from N = 500 to N = 1000 and from N = 1000 to N = 10,000.

(Intuitively, the number of synapses is equivalent to the number of "free parameters" in the classification model. As such, the more free parameters the model allows, the easier it is to find a separation between two sets of patterns.)

In conditions where the synapses were only on the soma or only on the basal tree, the performance of the BP is comparable to that of the M&P neuron (within ~10% classification accuracy for all values of N in the active, passive, and current conditions). In the condition where synapses were placed uniformly over the full tree, the discrepancies were somewhat larger (e.g. 13.7% for N = 1000 in the passive case; interestingly the performance improves when active mechanisms are added). However, when the synapses are all placed on the apical tuft of the biophysical cell (Figure 2B), the classification accuracy of the biophysical perceptron decreases dramatically, even in the presence of supra-linear boosting mechanisms such as NMDA receptors and active membrane ion channels. For example, in the condition with N = 1000 synapses, if the synapses are placed on the soma the neuron achieves 100% classification accuracy, whereas if the synapses are all placed on the apical tuft the neuron only achieves 67% accuracy. By switching from conductance synapses to current synapses in the apical tuft condition, however, it is possible to regain almost all of the "lost" classification accuracy (In the N = 1000 condition, from 67% with conductance synapses to 97% with current synapses).

We thus argue that the reason for the discrepancy in classification accuracy for the biophysical perceptron between the conditions wherein synapses are placed on the apical tuft as opposed to the soma or basal dendrites is due to the passive filtering properties of the neuronal cable. Specifically, the attenuation of voltage along the length of cable from apical tuft dendrites to the spike initiation zone means that the *effective weight* of that synapse - in other words the magnitude of the EPSP that it can produce in the soma - is greatly reduced. This phenomenon has been observed previously, but it has been argued (Häusser, 2001; Rumsey & Abbott, 2006) that the cell might be able to overcome this drop in voltage by simply increasing the strength (i.e. conductance) of distal synapses. We demonstrate, however, that this is not the case. We show (Figure 2C) that the perceptron learning algorithm will, on its own, increase the weights of apical tuft synapses far beyond the biologically plausible range of 0.2 - 1.3 nS (Eyal et al., 2018; Sarid, Bruno, Sakmann, Segev, & Feldmeyer, 2007) in its attempt to correctly classify all the patterns, but the classification accuracy of the apical tuft biophysical perceptron remains quite poor (see (Ilan, Gidon, & Segev, 2011) who show that the opposite phenomena will occur with a standard STDP rule, resulting in smaller synaptic conductances for distal synapses).

We claim that "democratization" via disproportionally increasing distal synaptic conductances does not solve the classification accuracy problem because effective synaptic weights are bounded by the synaptic reversal potential in the distal dendrites, even if one were to increase synaptic conductances to arbitrarily high values (Figure 3A). As such, the maximal effective synaptic weight (MESW) - defined as the peak somatic EPSP voltage when a given dendritic location attains the synaptic reversal potential - is equivalent to the synaptic driving force multiplied by the attenuation factor from that dendritic location to the soma. The MESWs for distal synapses are thus smaller than those for proximal synapses (Figure 3A). Because the apical tuft is electrotonically distant from the soma, synapses on the apical tuft have lower MESWs (Median for apical tuft: 1.0 mV, for average synapses in the full cell: 10.2 mV, and for synapses on the basal tree: 11.23 mv, Figure 3B). The fact that switching the apical synapses from conductance-based to current-based substantially improves classification accuracy supports the notion that voltage saturation due to synaptic reversal potential is responsible for the reduced performance of the apical tuft synapses.

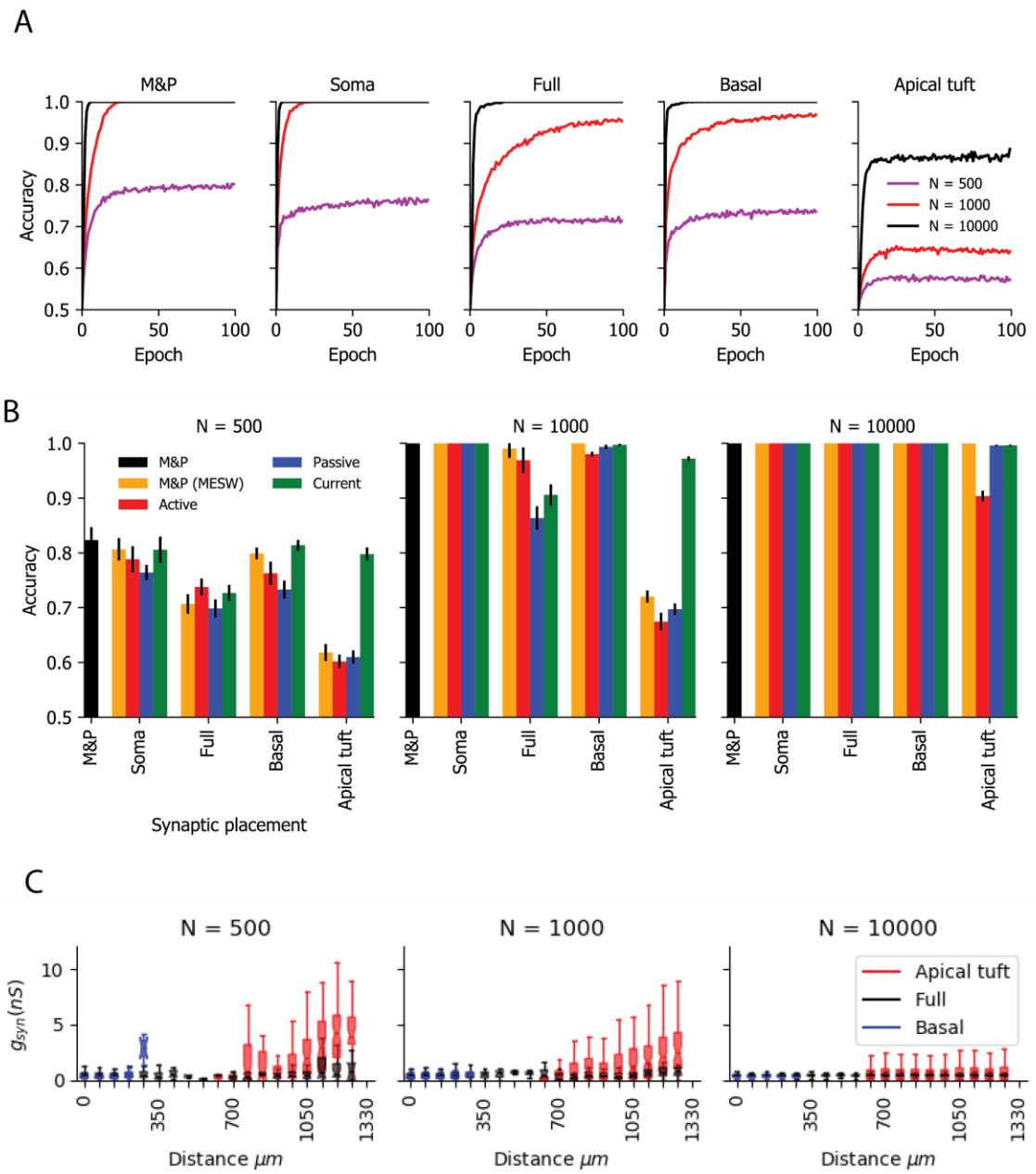


Figure 2. Learning the memorization task with the biophysical perceptron

(A) Learning curves for the memorization task, in which the neuron had to classify 1000 patterns (500 positive, 500 negative) in the fully active L5PC perceptron and in an M&P perceptron for different numbers of synapses (N, colored traces) and different conditions of synaptic placement in the L5PC model. Each trace is an average over ten trials. **Soma**: all N synapses are placed only at the soma, **Full**: all N synapses are uniformly distributed (per unit length) on the dendritic tree, **Basal**: all N synapses are uniformly distributed on basal tree (see Figure 1B) Apical tuft: all N synapses are uniformly distributed on the apical tuft (see Figure 1B).

(B) Accuracy for the memorization task for different models**.** Mean accuracy after 100 epochs for all excitation and synaptic placement conditions (error bars: standard deviation). The orange bar within each synaptic placement condition grouping shows the performance of a M&P neuron whose weights are constrained according to the distribution of maximum effective synaptic weights (MESW) of a respective biophysical neuron with cable filtering and conductance-based synapses (see **Methods** and Figure 3) for that synaptic placement condition. The black bar in each panel is the performance of an unconstrained M&P neuron with excitatory synapses. (See Supplementary Figure 1A for p-values).

(C) Value of synaptic conductances obtained after the completion of the learning algorithm as a function of distance of the synapses from the soma in 70-micron bins. The cases of synapses placed only on the apical tuft (red), basal tree (blue) and full placement conditions, black (as in panel A) are shown for a single exemplar run. Notches represent the median synaptic conductance for all synapses within that bin, box edges and error bars respectively represent the first and second quartiles of the data. Note the large synaptic conductance required to achieve the learning task for the case where the synapses are placed at the apical tuft.

From the standpoint of learning theory, the "cap" on the effective weights of distal apical synapses restricts the parameter space of the biophysical perceptron, reducing its capacity. To demonstrate this effect, we calculated the MESW for each synapse in the apical tuft and then imposed this distribution of MESWs onto an M&P perceptron. As predicted, the MESW-capped M&P perceptron produced a similar (reduced) classification capacity as does the biophysical perceptron when synapses were restricted to the apical tuft. We repeated this procedure for the other synaptic distribution conditions to demonstrate how MESWs affect the classification capacity of the basal tree, soma, and full neuron.

It should be noted that the limited capacity of the apical tuft is *not* due to the fact that apical synapses cannot induce the neuron to fire, because the neuron does appropriately fire in response to a subset of patterns after training. Moreover, if one only trains the apical tuft perceptron on "positive" patterns without forcing it not to spike in response to certain "negative" patterns, the neuron can indeed produce the appropriate spiking response. Indeed, a neuron with as few as 100 simultaneously active apical tuft synapses can correctly spike in response to 1000 "positive" learned patterns within a few epochs if it is not forced to reject any negative patterns. It is thus evident that the small classification capacity of the apical patterns is due to the restriction of the parameter space, not because the apical tuft has no ability to create a somatic spike.

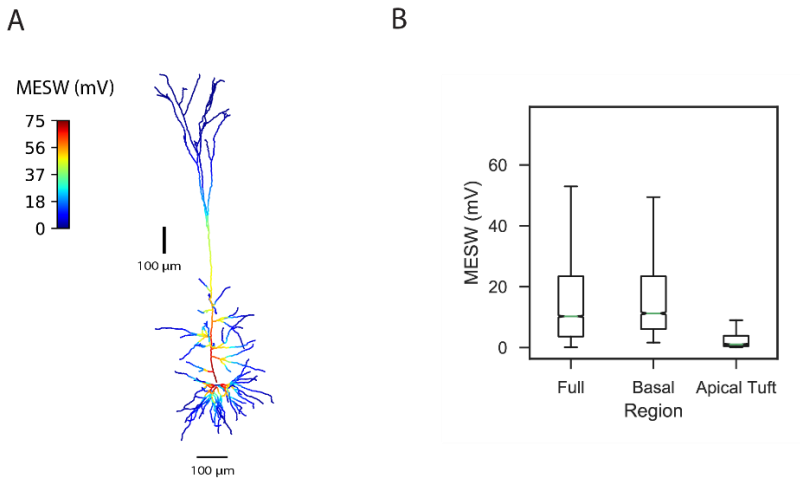


Figure 3. Effect of synapse location on the dendrite on the maximal effective synaptic weight (MESW).
(A) **A simulated L5 PC superimposed with the MESW values.** To calculate the MESWs for the cell, a transient AMPA/NMDA synapse with a massive peak conductance of 500 nS was placed at each dendritic segment of the neuron model, bringing the local synaptic voltage to the synaptic reversal potential of 0 mV (with resting potential of -76 mV). The MESW for a dendritic segment, shown as a heatmap, is defined as the peak depolarization obtained at the soma after 100 ms of synaptic activation. Note the steep voltage attenuation from distal dendritic branches to the soma (blue regions).
(B) Box-and-whiskers plot of MESWs when the synapses are placed on the apical tuft, basal tree, and the full dendritic tree. In these cases, 10,000 were uniformly distributed on the modelled tree and the MESW value was computed for each synapse (see **Methods)**. Notches represent the median MESWs for all synapses within that region, box edges and error bars respectively represent the first and second quartiles of the data.

Generalization task

To explore whether the apical tuft is always at a disadvantage when it comes to pattern classification, we tested the biophysical perceptron on a generalization task. Instead of "memorizing" a fixed set of patterns, in the generalization task the neuron was presented with presynaptic patterns of "stimulus-dependent" firing activity, which conveys information about which one of two stimuli, each drawn from two different classes, were being presented. A real-world example of this type of problem would be a neuron that needs to learn to discriminate between female and male faces. The visual features of any given female or male face - and thus the neural firing patterns of low-level sensory areas - will vary a great deal, but female and male faces still share certain within-group similarities. It is thus possible to characterize the presynaptic neural firing patterns as coming from two different probability distributions, P(pattern|female) and P(pattern/male), that a downstream neuron can distinguish between (Figure 4A). Importantly, these probability distributions can also take into account other sources of noise and variability in the brain, such as synaptic stochasticity or noisy inputs from other brain regions.

Our generalization task consisted of a simplified version of this problem, where the neuron had to distinguish between two multivariate Bernoulli distributions. We varied the difficulty of the discrimination task by drawing the multivariate Bernoulli parameters from a beta distribution with different variances each condition (see **Methods**). Drawing the Bernoulli parameters from high-variance beta distribution produced synaptic input patterns which were almost deterministic conditioned on the stimulus, drawing the Bernoulli parameters from low-variance distributions resulted in noisy synaptic input patterns that contained little information about the stimulus (Figure 4B). We were thus able to systematically test the performance of the biophysical perceptron on generalization tasks of inherently different difficulties (i.e. different signal-to-noise ratios in the input).

To train the neuron, we randomly generated binary input patterns drawn from the two probability distributions and presented 100 such patterns in each epoch, training the neuron in a supervised batch-learning protocol (see **Methods**). Because the patterns were randomly drawn at every iteration of the algorithm, it was unlikely that the neuron would see the same input pattern twice. Instead, the neuron would have to learn set its weights such that it would be able to learn to generalize and discriminate between the two probability distributions (e.g., between men and women), as is standard in most machine learning tasks. As expected, the neuron performed well on "easy" problems drawn from large-variance Beta distributions (100% accuracy with $\sigma^2 = 0.01$ in all conditions for N = 10000) and more poorly on problems drawn from low-variance Beta distributions (~70-85% accuracy for $\sigma^2 = 1e-5$ with N = 10000, Figure 5A-B).

Because the goal of the generalization task is qualitatively different than that of the memorization task, it is impossible to make direct comparison of their accuracy. We note, however, that while current synapses seemed to lead to slightly improved learning in all conditions, the difference between apical and soma classification capacity due to MESW discrepancies appears to be less relevant in this task as compared to the memorization task (Figure 2B). In the "hard" task where $N = 10000$, $\sigma^2 = 1e-5$,the difference in accuracy between the apical tree and the soma are small (soma: 72.8%, apical tuft: 70%). In the intermediate task, ($N = 10000, \sigma^2 = 5e-5$) the discrepancy was slightly larger (soma: 91.1%, apical tuft: 84.8%). In the easy task $\sigma^2 = 0.01$, both the soma and the apical tuft achieved perfect classification accuracy.

The discrepancy between the apical tuft and soma may be smaller in the generalization task relative to the memorization task because the memorization task is fundamentally about finding the correct weights to generate a hyperplane that will correctly separate between all relevant patterns, which requires a large amount of flexibility in the weight space for large numbers of patterns. By contrast, the generalization problem only contains two canonical "patterns" (i.e. probability distributions); the difficulty of the generalization problem is thus more related to the

intrinsic variability that emerges from the stochasticity of the input-output mapping. This can be seen in the weight distribution of the synapses after learning (Figure 5C) which maintain low values on the apical tuft even in the most difficult condition.

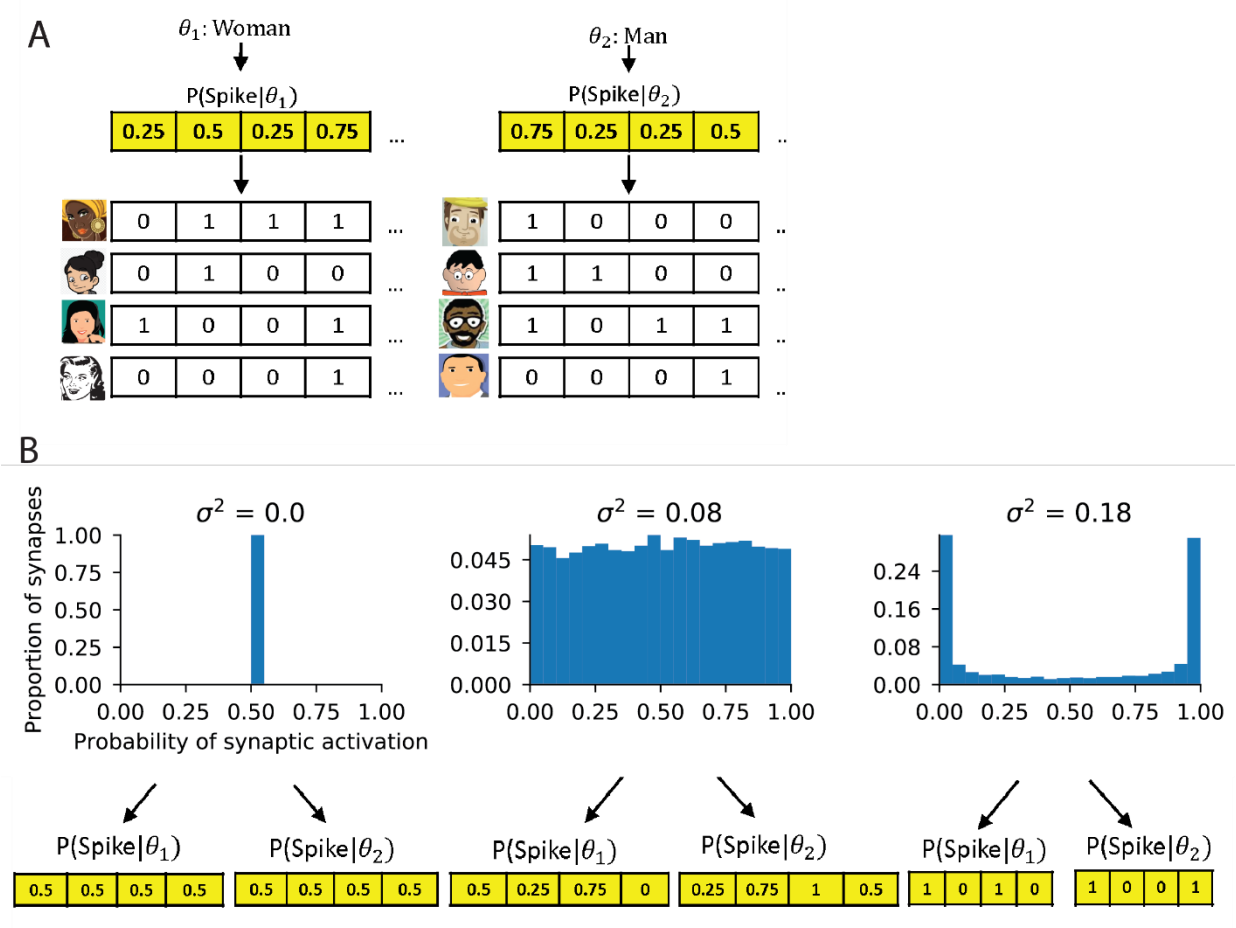


Figure 4. Generalization task

(A) Task schematic showing that the neuron needs to discriminate between two different classes of inputs, such as men (θ_1) and women (θ_2). Each class is defined by a unique multivariate Bernoulli probability distribution, with a fixed parameter for each synapse (yellow boxes). For the "Woman" class, a picture of a woman will elicit a presynaptic spike at the first synapse with a probability p = 0.25, at the second synapse with p = 0.5, and so on, whereas for the "Man" class, a picture of a man will elicit a presynaptic spike at the first synapse with a probability p = 0.75, at the second synapse with p = 0.25, and so on. Input patterns (i.e. individual men or women) are drawn from one of these two probability distributions and represented as a binary vector (white boxes) of presynaptic neurons either being active (1) or inactive (0). The biophysical perceptron needs to decide from which class the specific input pattern came from, by generating spikes for one class and no spikes for the other class. Unlike the memorization task, the patterns here are drawn randomly, so the neuron has to learn the underlying probability distribution of each input class. (Note: Our actual task consisted of binary patterns, not faces).
(B) Exemplar Bernoulli distributions drawn from Beta distributions. We vary the difficulty of the generalization task by drawing the parameters of the multivariate Bernoulli distribution (yellow boxes below, as in A) for each class from beta distributions (blue histograms) with different variances (σ^2). If we draw the Bernoulli parameters from beta distributions with a low variance ($\sigma^2 \approx 0$, left), the parameters for each synapse will be all be very close to the mean of the beta distribution. As such, each synapse conveys the same information, namely the overall mean firing probability for all synapses. If the overall mean firing probability for each class is the same, as in our case, it would be impossible to discern the distribution from which each pattern comes. If we draw the Bernoulli parameters from beta distributions with a large variance ($\sigma^2 \approx 0.18$, right), however, the Bernoulli parameters for each synapse will tends toward 0 and 1. So even if the overall average firing probability for both classes is identical, each synapse in a pattern can potentially provide a great deal of information about the class from which it comes. For example, if a particular synapse has a firing probability of 0 in θ_1 and a firing probability of 1 in θ_2, we can say on the basis of that single synapse whether any given pattern comes from θ_1 or θ_2. If we use a beta distribution with an intermediate variance ($\sigma^2 \approx 0.08$, center example) each synapse is only moderately informative about the class from which the pattern was drawn from, making for a classification problem of intermediate difficulty. (Note: histograms in image assume a symmetric beta distribution, where $\mu = 0.5$. This is for illustrative purposes only; the parameters for the distributions used in our generalization task are discussed in the text.)

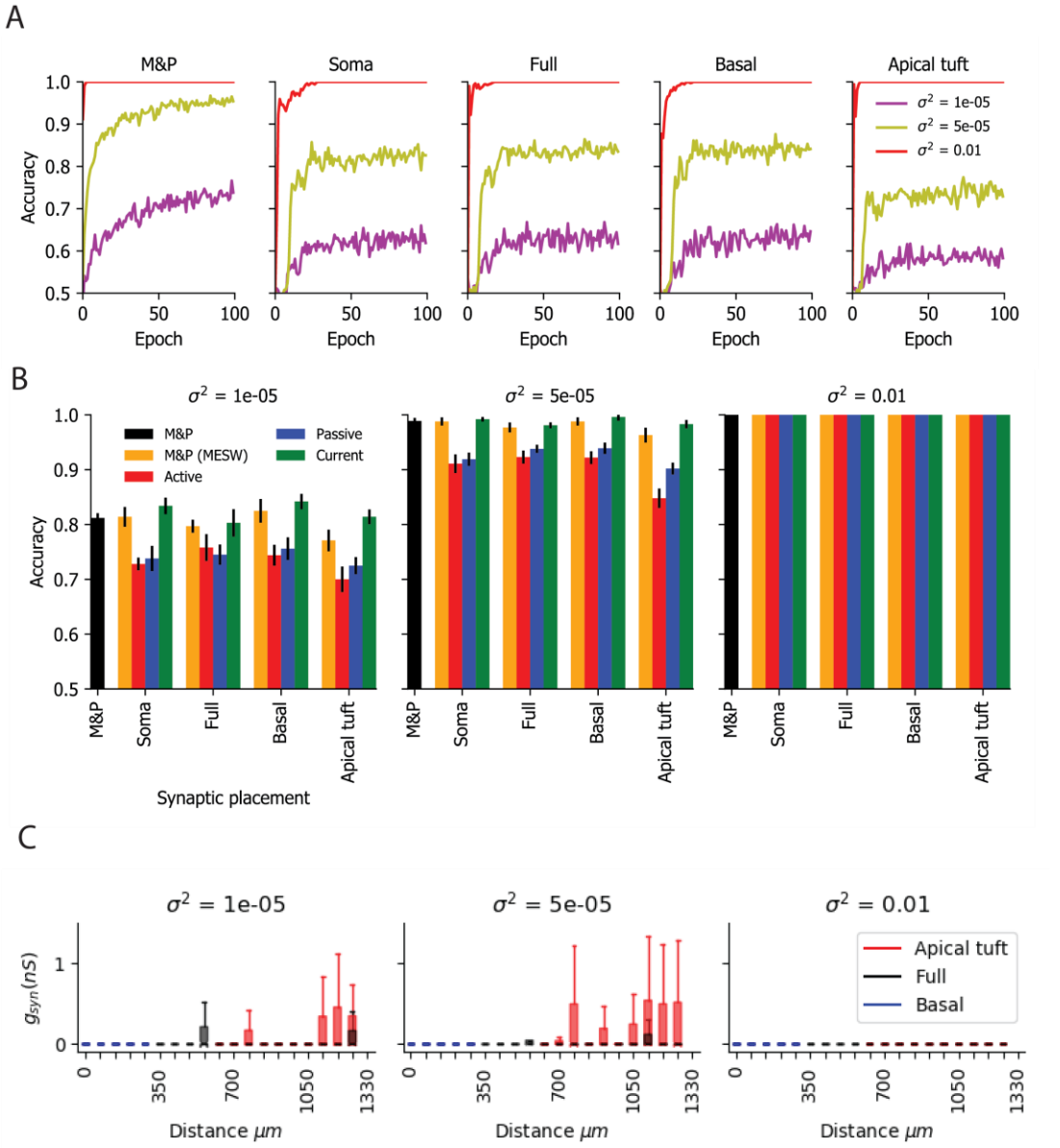


Figure 5. Learning the generalization task with the biophysical perceptron

(A) Learning curves for the generalization task, where the neuron had to classify patterns randomly drawn from one of two probability distributions, θ_1 and θ_2. Traces show the classification accuracy for the fully active L5PC perceptron and in an M&P perceptron for different values of the variance of the beta distribution (σ^2, see Figure 4B and **Methods**) and different conditions of synaptic placement in the L5PC. Each trace is an average over ten trials. Soma: all N synapses placed only at the soma, Full: all N synapses uniformly distributed (per unit length) on the dendritic tree, Basal: all N synapses uniformly distributed on basal tree (see Figure 1B) Apical tuft: all N synapses uniformly distributed on the apical tuft (see Figure 1B). Each epoch consisted of 100 patterns being presented to the neuron, 50 drawn from θ_1 and 50 drawn from θ_2. All conditions had 10,000 synapses, 200 of which were activated on average per pattern.

(B) Final accuracy for the memorization task**.** Mean accuracy (error bars: standard deviation) after 100 epochs for all excitation and synaptic placement conditions. The orange bar within each synaptic placement condition grouping shows the performance of a M&P neuron whose weights are constrained according to the distribution of maximum effective synaptic weights (MESW) of a respective biophysical neuron with cable filtering and conductance-based synapses (see **Methods** and Figure 3) for that synaptic placement condition. The black bar in each panel is the performance of an unconstrained M&P neuron with excitatory synapses. See Supplementary Figure 1B for p-values.

(C) Box-and-whisker plot for synaptic conductances obtained after the completion of the generalization task. Conductances are plotted as a function of distance from the soma in 70-micron bins for soma, apical tuft, and full placement conditions (as in panel A) from a single exemplar run. Notches represent the median synaptic conductance for all synapses within that bin, box edges and error bars respectively represent the first and second quartiles of the data.

Methods

We utilized a detailed biophysical model of a layer 5 pyramidal cell written in NEURON with a Python wrapper (Hines, Carnevale, & Hines, 2002). The parameters of the model are described in (Hay et al., 2011) Excitatory synapses were the AMPA/NMDA-based synapses as in (Muller and Reimann, 2011) with synaptic depression and facilitation parameters set to 0. In both experiments, for each number of synapses (N) that we tested, we placed all N synapses either on the soma, basal tree, or apical tuft according to a uniform spatial distribution. We tested our algorithm in three cases - a fully active tree with voltage-gated NMDA synapses and other dendritic nonlinearities, a passive tree with only AMPA synapses, and a passive tree where the synaptic input consisted of linear current injection as opposed to conductance-based synapses.

Memorization task

For the memorization task, each of the 1000 patterns was generated by randomly choosing 200 out of the N synapses to be activated. The patterns were then randomly assigned to either the positive or negative class. Patterns were presented to the cell by simultaneously stimulating the 200 active synapses with a single presynaptic spike at the beginning of the simulation. Simulations of the neuron were run with a Δt of 0.1 ms for a total of 100 ms. Patterns were considered to have been classified as "positive" if they produced at least one spike within the 100 ms time window and as "negative" if no spikes occurred.

We utilized an "online" version of the perceptron learning algorithm, applying the plasticity rule every time a pattern was presented to the neuron. Also, because we limited our analysis to excitatory synapses, we use the modified algorithm proposed in (Amit, Wong, & Campbell, 1999) for sign-constrained synapses, which ensures that synaptic weights never become negative.

The algorithm works as follows. An input pattern $\boldsymbol{x}$ is presented to the neuron, where $\boldsymbol{x}$ is a vector consisting of *N* binary inputs, each of which is labeled x_i and associated with a particular synapse on the dendritic tree with synaptic weight w_i. Each pattern has a target value, $y_0 \in \{1, -1\}$, associated with it, where 1 means "should spike" and -1 means "shouldn't spike." When the pattern is presented to the neuron via simultaneous activation all the synapses in the pattern, the soma of the neuron will produce a voltage response. If that voltage response contains at least one spike within 100 ms, we set the output variable $y = 1$. If the voltage response does not contain any spikes, we set $y = -1$. For each presynaptic input pattern, the plasticity rule for synapse i to update its weight w_i at time is defined as:

$$w_i \leftarrow \max(0, w_i + \eta dw_i) \tag{1}$$

Where dw_i is defined as:

$$dw_i = \begin{cases} 0, & y = y_0 \\ x_i, & y \neq y_0 \end{cases} \tag{2}$$

and η is the learning rate.

In other words, if the target output is the same as the actual output of the neuron, we do nothing. If the target is "should spike" and the neuron does not spike, we increase the weight of all synaptic inputs that were active in the pattern. If the target is "shouldn't spike" and the neuron does spike, we decrease the synaptic weights of all synaptic inputs that were active in the pattern, unless that would decrease the synaptic weight below 0, in which case we reduced the weight of that synapse to 0.

The accuracy of the neuron's output was calculated after each epoch, which consisted of a full pass of presenting every pattern (in random succession) to the neuron. To ensure that accuracy improved on every epoch and reached a reasonable asymptote for all conditions, we set the learning rate η to 0.002 for the condition with AMPA/NMDA synapses and an active tree, 0.005 for the condition with AMPA synapses and a passive tree, and 0.07 for the condition with current synapses and a passive tree. We also used the "momentum" technique (Rumelhart et al., 1986) to improve learning speed.

M&P model (not constrained by synaptic battery)

To compare the BP to an equivalent M&P perceptron (Figures 2B and 5B, black bars) we used a M&P perceptron with only excitatory weights as described in (Amit et al., 1999) (See Equation (1) and (2) above). A M&P neuron with no inputs would have a "bias" input value of -77.13 to mimic the resting potential of the BP and a "spiking threshold" of -22.8 to mimic the voltage spiking threshold of a real neuron. We used a learning rate η of 0.01 which was dynamically modified in the learning algorithm via the momentum technique (Rumelhart et al., 1986).

MESW calculation and MESW-constrained M&P model

To calculate the MESWs for the L5PC model, we placed a transient AMPA/NMDA synapse (see **Methods**) with massive synaptic conductance (500 nS) at each segment of the neuron model, bringing the local synaptic voltage to synaptic reversal potential of 0 mV (with resting potential of -76 mV). The MESW for a dendritic segment is defined as the peak depolarization obtained at the soma within 100 ms after synaptic activation. (Figure 3A). To create an MESW-constrained M&P models for different synaptic placement conditions, we uniformly distributed 10,000 AMPA/NMDA either over the full dendritic tree, only on the apical tuft, or only on the basal tree. We then calculated the MESW for each synapse and calculated the probability distribution for an synapse randomly placed on the apical tuft, full dendritic tree, or basal tree would have a given MESW. (The median and quartile values of these distributions are shown in the box-and-whisker plot in Figure 3b.) We then created M&P neurons where each synapse was individually given a "cap" drawn randomly from the probability distribution which would prevent the weight of that input from increasing above a certain value. In other words, if the plasticity algorithm (Equation 1) would bring w_i to be greater than cap, c_i, we would "freeze" the weight at c_i. Formally, this means that the plasticity algorithm for the MESW-capped neuron is

$$w_i \leftarrow \max\left(0, min(c_i, w_i + \eta dw_i)\right) \quad (3)$$

Where η and dw_i are as defined above in Equation (2).

Generalization task

In the second task (generalization), we generated two multivariate Bernoulli distributions in the following manner. For each of the two distributions, θ_1 and θ_2, The Bernoulli parameter of each synapse was independently drawn from a beta distribution to produce firing probabilities that varied between 0 and 1. A beta distribution is usually characterized in terms of the parameters α and β:

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad (4)$$

Where $\alpha, \beta > 0$, $x \in [0,1]$, and $\Gamma(z)$ is the gamma function. We reparameterized the beta distribution in terms of the mean (μ) and the variance (σ^2) instead of the traditional α and β by substituting the following values for α and β (Max, 2016):

$$\alpha = \left(\frac{1-\mu}{\sigma^2} - \frac{1}{\mu}\right)\mu^2 \quad (5)$$

$$\beta = \alpha\left(\frac{1}{\mu} - 1\right) \quad (6)$$

Where

$$\mu \in 0,1 \quad (7)$$

And

$$\sigma^2 \leq \mu(1-\mu) \quad (8)$$

This enabled us to fix the average firing rate for each pattern to 200 active synapses by setting $\mu = 200/N$. We then chose different values for the variance of the beta distribution ($\sigma^2 \in \{1e-5, 5e-5, 0.01\}$) to demonstrate the efficacy of the algorithm in discriminating between probability distributions with different levels of difficulty. We fixed N at 10,000 synapses for all conditions in the generalization task. Patterns drawn from a Bernoulli distribution with parameters generated by a high-variance beta distribution tend to be similar to each other, because the Bernoulli parameters tend to be close to 0 and 1. Patterns drawn from a Bernoulli distribution with parameters generated by a low-variance beta distribution, however, will tend to be different from each other, because the Bernoulli parameters for all of the synapses will tend to be close to the overall mean firing rate. In other words, when σ^2 is large, there is a high cohesiveness among patterns in the same class, making the classes easy to distinguish from each other, whereas when σ^2 is small, there is lower within-class cohesiveness making the classification problem harder. (Another way to think about this is that when σ^2 is large and the Bernoulli parameters are closer to 0 or 1, there is more mutual information per synapse between the pattern and the probability distribution from which it is drawn, because the pattern is almost fully determined by the probability distribution.)

In every epoch of the learning task, we presented the neuron with 50 patterns randomly drawn from θ_1 and 50 patterns drawn randomly from θ_2 for a total of 100 patterns per epoch (the order of the presentation of patterns from θ_1 and θ_2 was also randomized). Unlike the memorization task, the generalization used a batch learning protocol for the perceptron learning rule, updating the synapses after every epoch as opposed to every iteration.

Simulations were all performed using Neuron v.7.3 (Hines, Carnevale, & Hines, 2002) running on a multi-core cluster computer. An average simulation time for a complete run of the learning algorithm (i.e. 100 epochs) was several days, depending on the task.

Discussion

In the simulations described above, we have demonstrated that the perceptron learning algorithm can indeed be implemented in a detailed biophysical model of L5 pyramidal cell with conductance-based synapses, active dendrites. This is despite the fact that the perceptron learning algorithm traditionally assumes a cell which integrates its inputs linearly, which is not the case for detailed biophysical neurons with a variety of nonlinear active and passive properties, and conductance-based synapses. Evidently, although such nonlinearities are present, they do not qualitatively prevent the cell from learning in a perceptron-like manner (the memorization task, Figures 1-2), nor do they enhance the learning ability of the cell (at least if they are not explicitly utilized in the algorithm). That being said, the ability of a biophysical perceptron to distinguish between different patterns of excitatory synaptic input does depend on the location of the relevant synapses. Specifically, if all the synapses are located proximally to the soma, such as on the proximal basal tree and apical obliques, the cell has a classification capacity similar to that of the M&P perceptron. However, for patterns consisting of more distal synapses, such as those on the apical tuft, the classification capacity of the BP is reduced. We showed that this is due to the reduced effectiveness of distal synapses, which limits the parameter space of the learning algorithm and thus hampers classification capacity.

We demonstrated that the issue of diminished classification capacity in the apical tuft is negligible in a generalization task. This indicates that, while the maximum effective synaptic weights of neuron may be somewhat limiting for a neuron's classification accuracy in the memorization task, they do not hamper the apical tuft's ability to independently function reasonably well as a perceptron for a biologically plausible stochastic generalization task (Figure 4-5).

The above discussion of pyramidal cells considers that the cell separately classifies inputs that synapse onto different regions of the cell (such as the apical tuft and the basal tree) and doesn't necessarily simultaneously integrate all the synaptic input impinging on the cell. This decision was motivated by a growing evidence that different parts of the dendritic tree may play a key role in shaping the neuron's output. From anatomical studies, it is known that axons from different brain regions preferentially synapse onto particular regions of layer 5 pyramidal cells. For example, basal dendrites tend to receive local inputs whereas apical tuft dendrites receive long-range cortical inputs (Budd, 1998; Crick & Asanuma, 1986; Spratling, 2002; Spruston, 2008). This has led to theories of neuronal integration for layer 5 pyramidal cells that involve a "bottom-up" stream of information entering the basal dendrites and "top-down" signals coming to the apical tuft (M. Larkum, 2013; Manita et al., 2015; Siegel, Körding, & König, 2000). Moreover, it has recently been shown experimentally that when experiencing somatosensory stimulation, layer 5 pyramidal cells in S1 first exhibit an increase in firing rate corresponding to the bottom-up sensory input (ostensibly to the basal tree), and then, 30 ms later, receive top-down input to the apical tuft from M2 (Manita et al., 2015). This indicates the presence of temporally segregated time windows in which the cell separately integrates input from the apical and basal tree. There is also work suggesting that plasticity rules may function differently in different regions of the cell (Gordon, Polsky, & Schiller, 2006), again indicating that different regions of the cell might serve as input regions to distinct information pathways, and, as such, may have different priorities underlying the decision of when the cell will or will not fire. Taken together, the above studies strongly suggest that the apical tuft and basal dendrites can and should be studied as independent integration units.

Our study made several simplifications in the learning and plasticity processes used as compared to those found in biology. Critically, our plasticity algorithm utilized only excitatory synapses and did not consider the effect of inhibition on learning. This is not because we believe that inhibition plays no role in learning; on the contrary, inhibitory synapses are essential both for the learning process and in defining the input-output function of the cell (Kullmann, Moreau, Bakiri,

& Nicholson, 2012; Müllner, Wierenga, & Bonhoeffer, 2015; Wulff et al., 2009). However, by restricting ourselves to excitatory synapses, we were able to isolate important biophysical properties of excitatory synapses – namely the impact of synaptic saturation (the MESWs) that might have been masked in the presence of inhibition. The focus on excitatory synapses also enables our work to be directly compared to studies of excitatory perceptron-like learning done on Purkinje cells, such as the work of (Brunel, Hakim, Isope, Nadal, & Barbour, 2004; Steuber et al., 2007). Future work on the "biophysical perceptron" will include the role of inhibitory synapses; in this case special care must be taken to understand how inhibitory inputs interact with excitatory inputs on different locations of the cell (Doron, Chindemi, Eilif Muller, & Segev, 2017; Gidon & Segev, 2012). The addition of synaptic inhibition has the potential to increase the classification capacity of the cell (Chapeton et al., 2012), and localized inhibition may allow for additional forms of compartmentalized computation at the single neuron level.

Our focus on perceptron-like learning constitutes an additional simplification, as perceptron learning ignores the potential contributions that dendritic nonlinearities such as local NMDA spikes (Polsky, Mel, & Schiller, 2004; Schiller, Major, Koester, & Schiller, 2000), dendritic Na^+ spikes (Golding & Spruston, 1998; Sun, Srinivas, Sotayo, & Siegelbaum, 2014), dendritic Ca^{2+} spikes (Cichon & Gan, 2015; Kampa, Letzkus, & Stuart, 2006; Magee & Johnston, 1995) may impact learning in classification tasks. Although a variety of dendritic nonlinearities are present in our model, we did not make explicit use of them in our plasticity rule. Indeed, some models of dendritic integration such as the Clusteron (Mel, 1991, 1992) and the two-layer model (Poirazi & Mel, 2001) treat the NMDA spike as critical for dendritic computation. In particular, these models treat clustering of nearby synapses, and "structural plasticity", or the relocation of synaptic inputs within and between branches as crucial for learning (Kastellakis, Cai, Mednick, Silva, & Poirazi, 2015; M. E. Larkum & Nevian, 2008; Losonczy, Makara, & Magee, 2008; Mel, Schiller, & Poirazi, 2017; Trachtenberg et al., 2002; Weber et al., 2016). The present study did not address the role of synaptic clustering in learning, a promising future direction would be to combine the weight-based learning rules used in our study with the structural plasticity algorithm as discussed in (Mel, 1992).

Another crucial element that remains to be studied in detailed biophysical models is the role of the timing of both the input and output of pyramidal cells in learning and computation. Regarding input timing, some theoretical work has been done on the M&P perceptron, which has been extended in a variety of ways to take into account several components of real neurons, like the Tempotron, which uses a leaky integrate and fire mechanism (Gütig & Sompolinsky, 2006) and conductance-based (rather than current-based) synapses (Gütig & Sompolinsky, 2009) to classify spatiotemporal input patterns. Regarding output timing and firing rate, learning rules like the one from (Gutig, 2016) can learn to solve the temporal credit-assignment by producing different spike rates for different inputs. Similarly, the Chronotron (Florian, 2012) considers learning rules that generate precisely timed output spikes. It is not clear to what extent these particular plasticity algorithms are truly "biological", but there is no question that temporal sequence learning is an essential feature of the brain (Aslin, Saffran, & Newport, 1998; Moldwin, Schwartz, & Sussman, 2017; Xu, Jiang, Poo, & Dan, 2012). The addition of a temporal dimension increases the classification capacity of the cell, as discussed in (Gütig & Sompolinsky, 2006).

The present study shows that, by implementing the perceptron learning rule, layer 5 cortical pyramidal cells are powerful learning and generalization units, comparable – at the very least -- to the abstract M&P perceptron. Other plasticity rules which take into account synaptic clustering, input and output timing, and integration between the apical and basal regions of pyramidal cells will be explored in further studies in detailed biophysical models to determine their biological plausibility and classification capacity. Until then, our study should be viewed as a baseline for comparison of any future work implementing learning algorithms in detailed biophysical models of neurons.

References

Albus, J. S. (1971). A Theory of Cerebellar Function. *Mathematical Biosciences*, *10*, 25–61.

Amit, D. J., Wong, K. Y. M., & Campbell, C. (1999). Perceptron learning with sign-constrained weights. *Journal of Physics A: Mathematical and General*, *22*(12), 2039–2045. https://doi.org/10.1088/0305-4470/22/12/009

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics By 8-Month-Old Infants, *9*(4), 321–324.

Brunel, N., Hakim, V., Isope, P., Nadal, J. P., & Barbour, B. (2004). Optimal information storage and the distribution of synaptic weights: Perceptron versus Purkinje cell. *Neuron*, *43*(5), 745–757. https://doi.org/10.1016/j.neuron.2004.08.023

Budd, J. M. L. (1998). Extrastriate feedback to primary visual cortex in primates: A quantitative analysis of connectivity. *Proceedings of the Royal Society B: Biological Sciences*, *265*(1400), 1037–1044. https://doi.org/10.1098/rspb.1998.0396

Chapeton, J., Fares, T., LaSota, D., & Stepanyants, A. (2012). Efficient associative memory storage in cortical circuits of inhibitory and excitatory neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(51), E3614-22. https://doi.org/10.1073/pnas.1211467109

Cichon, J., & Gan, W.-B. (2015). Branch-specific dendritic Ca2+ spikes cause persistent synaptic plasticity. *Nature*, *520*(7546), 180–185. https://doi.org/10.1038/nature14251

Crick, F. ;, & Asanuma, C. (1986). Certain aspects of the anatomy and physiology of the cerebral cortex. In and T. P. R. G. In Rumelhart, D. E., McClelland, J. L. (Ed.), *Parallel Distributed Processing: Explorations in the Microstructures of Cognition. Volume 2: Psychological and Biological Models* (pp. 333–71). MIT Press, Cambridge, MA.

Doron, M., Chindemi, G., Eilif Muller, H. M., & Segev, I. (2017). Timed Synaptic Inhibition Shapes NMDA Spikes, Influencing Local Dendritic Processing and Global I/O Properties of Cortical Neurons. *Cell Reports*, *21*, 1550–1561. https://doi.org/10.1016/j.celrep.2017.10.035

Eilif Muller, Michael Reimann, S. R. (n.d.). Modification of ProbAMPANMDA: 2-State model. Retrieved from https://github.com/BlueBrain/BluePyOpt/blob/master/examples/l5pc/mechanisms/ProbAMPANMDA_EMS.mod

Eyal, G., Verhoog, M. B., Testa-Silva, G., Deitcher, Y., Benavides-Piccione, R., DeFelipe, J., … Segev, I. (2018). Human Cortical Pyramidal Neurons: From Spines to Spikes via Models. *Frontiers in Cellular Neuroscience*, *12*(June), 1–24. https://doi.org/10.3389/fncel.2018.00181

Florian, R. V. (2012). The chronotron: A neuron that learns to fire temporally precise spike patterns. *PLoS ONE*, *7*(8). https://doi.org/10.1371/journal.pone.0040233

Gidon, A., & Segev, I. (2012). Principles Governing the Operation of Synaptic Inhibition in Dendrites. *Neuron*, *75*(2), 330–341. https://doi.org/10.1016/j.neuron.2012.05.015

Golding, N. L., & Spruston, N. (1998). Dendritic sodium spikes are variable triggers of axonal action potentials in hippocampal CA1 pyramidal neurons. *Neuron*, *21*(5), 1189–1200. https://doi.org/10.1016/S0896-6273(00)80635-2

Gordon, U., Polsky, A., & Schiller, J. (2006). Plasticity Compartments in Basal Dendrites of Neocortical Pyramidal Neurons. *Journal of Neuroscience*, *26*(49), 12717–12726. https://doi.org/10.1523/JNEUROSCI.3502-06.2006

Gutig, R. (2016). Spiking neurons can discover predictive features by aggregate-label learning. *Science*, *351*(6277), aab4113-aab4113. https://doi.org/10.1126/science.aab4113

Gütig, R., & Sompolinsky, H. (2006). The tempotron: a neuron that learns spike timing-based decisions. *Nature Neuroscience*, *9*(3), 420–428. https://doi.org/10.1038/nn1643

Gütig, R., & Sompolinsky, H. (2009). Time-warp-invariant neuronal processing. *PLoS Biology*,

7(7). https://doi.org/10.1371/journal.pbio.1000141

Häusser, M. (2001). Synaptic function: Dendritic democracy. *Current Biology*, *11*(1), 10–12. https://doi.org/10.1016/S0960-9822(00)00034-8

Hay, E., Hill, S., Schürmann, F., Markram, H., & Segev, I. (2011). Models of neocortical layer 5b pyramidal cells capturing a wide range of dendritic and perisomatic active properties. *PLoS Computational Biology*, *7*(7). https://doi.org/10.1371/journal.pcbi.1002107

Hines, M. L., Carnevale, N. T., & Hines, M. L. (2002). The NEURON Simulation Environment telephone 203−737−4232 fax 203−785−6990, 1–9. Retrieved from http://mitpress.mit.edu

Ilan, L. B., Gidon, A., & Segev, I. (2011). Signals, 989–998. https://doi.org/10.1152/jn.00612.2010.

Kampa, B. M., Letzkus, J. J., & Stuart, G. J. (2006). Requirement of dendritic calcium spikes for induction of spike-timing-dependent synaptic plasticity. *Journal of Physiology*, *574*(1), 283–290. https://doi.org/10.1113/jphysiol.2006.111062

Kastellakis, G., Cai, D. J., Mednick, S. C., Silva, A. J., & Poirazi, P. (2015). Synaptic clustering within dendrites: An emerging theory of memory formation. *Progress in Neurobiology*, *126*, 19–35. https://doi.org/10.1016/j.pneurobio.2014.12.002

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, 1–9. https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007

Kullmann, D. M., Moreau, A. W., Bakiri, Y., & Nicholson, E. (2012). Plasticity of Inhibition. *Neuron*, *75*(6), 951–962. https://doi.org/10.1016/j.neuron.2012.07.030

Larkum, M. (2013). A cellular mechanism for cortical associations: An organizing principle for the cerebral cortex. *Trends in Neurosciences*, *36*(3), 141–151. https://doi.org/10.1016/j.tins.2012.11.006

Larkum, M. E., & Nevian, T. (2008). Synaptic clustering by dendritic signalling mechanisms. *Current Opinion in Neurobiology*, *18*(3), 321–331. https://doi.org/10.1016/j.conb.2008.08.013

Losonczy, A., Makara, J. K., & Magee, J. C. (2008). Compartmentalized dendritic plasticity and input feature storage in neurons. *Nature*, *452*(7186), 436–441. https://doi.org/10.1038/nature06725

Magee, J. C., & Johnston, D. (1995). Characterization of Single Voltage-Gated Na+ and Ca2+ Channels in Apical Dendrites of Rat {CA1} Pyramidal Neurons. *J. Physiol.*, *481*, 67–90.

Manita, S., Suzuki, T., Homma, C., Matsumoto, T., Odagawa, M., Yamada, K., … Murayama, M. (2015). A Top-Down Cortical Circuit for Accurate Sensory Perception. *Neuron*, *86*(5), 1304–1316. https://doi.org/10.1016/j.neuron.2015.05.006

Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. *Science*, *275*(5297), 213–215. https://doi.org/10.1126/science.275.5297.213

Marr, D. (1969). A THEORY OF CEREBELLAR CORTEX. *J. Physiol*, *202*, 437–470. https://doi.org/10.2307/1776957

Max. (n.d.). Calculating the parameters of a beta distribution using the mean and variance. Retrieved from http://stats.stackexchange.com/questions/12232/calculating-the-parameters-of-a-beta-distribution-using-the-mean-and-variance

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133. https://doi.org/10.1007/BF02478259

Megías, M., Emri, Z., Freund, T. F., & Gulyás, a I. (2001). Total number and distribution of inhibitory and excitatory synapses on hippocampal CA1 pyramidal cells. *Neuroscience*, *102*(3), 527–540. https://doi.org/S0306-4522(00)00496-6 [pii]

Mel, B. W. (1991). The clusteron: Toward a simple abstraction for a complex neuron. *Nips*, 35–42.

Mel, B. W. (1992). NMDA-Based Pattern Discrimination in a Modeled Cortical Neuron. *Neural Computation*, *4*, 502–517. https://doi.org/10.1162/neco.1992.4.4.502

Mel, B. W., Schiller, J., & Poirazi, P. (2017). Synaptic plasticity in dendrites: complications and coping strategies. *Current Opinion in Neurobiology*, *43*, 177–186. https://doi.org/10.1016/j.conb.2017.03.012

Moldwin, T., Schwartz, O., & Sussman, E. S. (2017). Statistical Learning of Melodic Patterns Influences the Brain's Response to Wrong Notes. *Journal of Cognitive Neuroscience*, *29*(12). https://doi.org/10.1162/jocn_a_01181

Müllner, F. E., Wierenga, C. J., & Bonhoeffer, T. (2015). Precision of Inhibition: Dendritic Inhibition by Individual GABAergic Synapses on Hippocampal Pyramidal Cells Is Confined in Space and Time. *Neuron*, *87*(3), 576–589. https://doi.org/10.1016/j.neuron.2015.07.003

Poirazi, P., Brannon, T., & Mel, B. W. (2003). Pyramidal neuron as two-layer neural network. *Neuron*, *37*(6), 989–999. https://doi.org/10.1016/S0896-6273(03)00149-1

Poirazi, P., & Mel, B. W. (2001). Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. *Neuron*, *29*(3), 779–796. https://doi.org/10.1016/S0896-6273(01)00252-5

Polsky, A., Mel, B. W., & Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. *Nature Neuroscience*, *7*(6), 621–627. https://doi.org/10.1038/nn1253

Quiroga, R. Q., Kreiman, G., Koch, C., & Fried, I. (2008). Sparse but not "Grandmother-cell" coding in the medial temporal lobe. *Trends in Cognitive Sciences*, *12*(3), 87–91. https://doi.org/10.1016/j.tics.2007.12.003

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(7045), 1102–1107. https://doi.org/10.1038/nature03687

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in …. *Psychological Review*, *65*(6), 386–408. https://doi.org/10.1037/h0042519

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Rumsey, C. C., & Abbott, L. F. (2006). Synaptic Democracy in Active Dendrites. *Journal of Neurophysiology*, *96*(5), 2307–2318. https://doi.org/10.1152/jn.00149.2006

Safaryan, K., Maex, R., Davey, N., Adams, R., & Steuber, V. (2017). Nonspecific synaptic plasticity improves the recognition of sparse patterns degraded by local noise. *Scientific Reports*, *7*(October 2016), 46550. https://doi.org/10.1038/srep46550

Sarid, L., Bruno, R., Sakmann, B., Segev, I., & Feldmeyer, D. (2007). Modeling a layer 4-to-layer 2/3 module of a single column in rat neocortex: Interweaving in vitro and in vivo experimental observations. *Proceedings of the National Academy of Sciences*, *104*(41), 16353–16358. https://doi.org/10.1073/pnas.0707853104

Schiller, J., Major, G., Koester, H. J., & Schiller, Y. (2000). NMDA spikes in basal dendrites of cortical pyramidal neurons. *Nature*, *404*(6775), 285–289. https://doi.org/10.1038/35005094

Siegel, M., Körding, K. P., & König, P. (2000). Integrating top-down and bottom-up sensory processing by somato-dendritic interactions. *Journal of Computational Neuroscience*, *8*(2), 161–173. https://doi.org/10.1023/A:1008973215925

Spratling, M. W. (2002). Cortical region interactions and the functional role of apical dendrites. *Behavioral and Cognitive Neuroscience Reviews*, *1*(3), 219–228. https://doi.org/10.1177/1534582302001003003

Spruston, N. (2008). Pyramidal neurons: Dendritic structure and synaptic integration. *Nature Reviews Neuroscience*, *9*(3), 206–221. https://doi.org/10.1038/nrn2286

Steuber, V., Mittmann, W., Hoebeek, F. E., Silver, R. A., De Zeeuw, C. I., Häusser, M., & De Schutter, E. (2007). Cerebellar LTD and Pattern Recognition by Purkinje Cells. *Neuron*, *54*(1), 121–136. https://doi.org/10.1016/j.neuron.2007.03.015

Sun, Q., Srinivas, K. V, Sotayo, A., & Siegelbaum, S. a. (2014). Dendritic Na(+) spikes enable cortical input to drive action potential output from hippocampal CA2 pyramidal neurons. *ELife*, *3*, 1–24. https://doi.org/10.7554/eLife.04551

Trachtenberg, J. T., Chen, B. E., Knott, G. W., Feng, G., Sanes, J. R., Egbert Welker, & Karel Svoboda. (2002). Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature*, *420*(19), 789–794. https://doi.org/10.1111/j.1528-1157.1986.tb03495.x

Weber, J. P., Andrásfalvy, B. K., Polito, M., Magó, Á., Ujfalussy, B. B., & Makara, J. K. (2016). Location-dependent synaptic plasticity rules by dendritic spine cooperativity. *Nature Communications*, *7*, 11380. https://doi.org/10.1038/ncomms11380

Wulff, P., Schonewille, M., Renzi, M., Viltono, L., Sassoè-Pognetto, M., Badura, A., … De Zeeuw, C. I. (2009). Synaptic inhibition of Purkinje cells mediates consolidation of vestibulo-cerebellar motor learning. *Nature Neuroscience*, *12*(8), 1042–1049. https://doi.org/10.1038/nn.2348

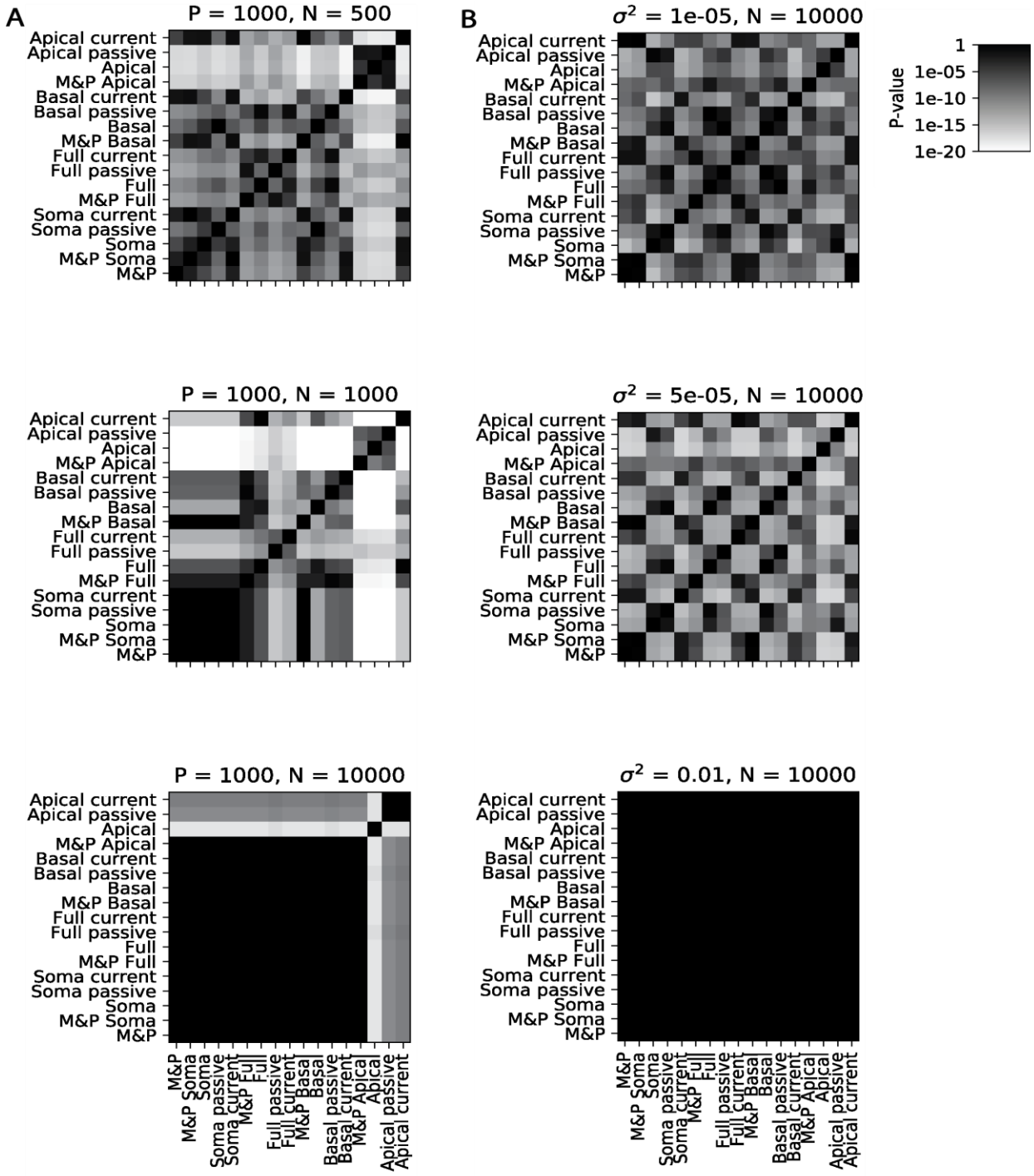
Xu, S., Jiang, W., Poo, M., & Dan, Y. (2012). Activity recall in a visual cortical ensemble. *Nature Neuroscience*, *15*(3), 449–455. https://doi.org/10.1038/nn.3036

Zilberter, M., Holmgren, C., Shemer, I., Silberberg, G., Grillner, S., Harkany, T., & Zilberter, Y. (2009). Input specificity and dependence of spike timing-dependent plasticity on preceding postsynaptic activity at unitary connections between neocortical layer 2/3 pyramidal cells. *Cerebral Cortex*, *19*(10), 2308–2320. https://doi.org/10.1093/cercor/bhn247

Acknowledgements

We would like to thank Oren Amsalem for his assistance in many aspects of this work, particularly his help with the use of the NEURON software. We also would like to thank David Beniaguev, Guy Eyal, and Michael Doron for their insightful discussions about machine learning and neuronal biophysics. Additionally, we appreciate Nizar Abed's support in maintaining the computer systems used to perform our simulations and data analysis. This work was supported by the EU Horizon 2020 program (720270, Human Brain Project), a grant from Huawei Technologies Co., Ltd., and by a grant from the Gatsby Charitable Foundation.

Supplementary Figures



Supplementary Figure 1. Significance tests

(A) P-values obtained in memorization task by pairwise T-tests between each experimental condition for different values of N (see Figure 2B).
(B) P-values for generalization task. P-values obtained in generalization task by pairwise T-tests between each experimental condition for different values of σ^2 (See Figure 5B).