Multiple Synaptic Contacts combined with Dendritic Filtering enhance Spatio-Temporal Pattern Recognition capabilities of Single Neurons

David Beniaguev[1], Sapir Shapira[1], Idan Segev[1,2] and Michael London[1,2]

[1]The Edmond and Lily Safra Center for Brain Sciences (ELSC), The Hebrew University of Jerusalem, Israel

[2]Department of Neurobiology, The Hebrew University of Jerusalem, Israel.

Abstract

A cortical neuron typically makes multiple synaptic contacts on the dendrites of a post-synaptic target neuron. The functional implications of this apparent redundancy are unclear. The dendritic location of a synaptic contact affects the time-course of the somatic post-synaptic potential (PSP) due to dendritic cable filtering. Consequently, a single pre-synaptic axonal spike results with a PSP composed of multiple temporal profiles. Here, we developed a "filter-and-fire" (F&F) neuron model that captures these features and show that the memory capacity of this neuron is threefold larger than that of a leaky integrate-and-fire (I&F) neuron, when trained to emit precisely timed output spikes for specific input patterns. Furthermore, the F&F neuron can learn to recognize spatio-temporal input patterns, e.g., MNIST digits, where the I&F model completely fails. Multiple synaptic contacts between pairs of cortical neurons are therefore an important feature rather than a bug and can serve to reduce axonal wiring requirements.

Keywords: Synaptic Integration; Dendritic Computation; Cable Theory; Compartmental Models; Neural Code; Computational Neuroscience; Biological Learning; Machine learning.

Introduction

Neurons in the central nervous system (CNS) connect to each other via chemical synapses. In recent decades it was found that two neurons that are synaptically connected typically connect via multiple synaptic contacts rather than a single contact (Holler et al. 2021; Silver et al. 2003; Feldmeyer, Lübke, and Sakmann 2006; Shepherd et al. 2005; Markram et al. 1997). Multiple synaptic contacts that originate from a single pre-synaptic axon often impinge on different parts of the dendritic tree of the post-synaptic neuron (Feldmeyer, Lübke, and Sakmann 2006; Silver et al. 2003; Holler et al. 2021). Furthermore, if contacts were just based on "Peters' rule" (Peters and Feldman 1976) (namely by axon-dendrite proximity), then one would expect that the distribution of the number of multiple contacts would be exponential, with one contact per axon being the most frequent case (Fares and Stepanyants 2009; Markram et al. 2015; Rees, Moradi, and Ascoli 2017) which is far from what was empirically observed. The deviation from the distribution predicted by Peters' rule suggests that the number of synaptic contacts between two connected neurons is tightly controlled and is thus likely to serve for a functional purpose.

Several phenomenological models have attempted to explain how multiple synaptic contacts between the pre-synaptic and post-synaptic neurons are formed (Fares and Stepanyants 2009), but very few studies have tried to tackle the question of how might they be beneficial from a computational perspective, but see (Sezener et al. 2021; Camp, Mandivarapu, and Estrada 2020; Jones and Kording 2021; Zhang, Hu, and Liu 2020; Hiratani and Fukai 2018; Acharya et al. 2021). It is typically thought that this redundancy overcomes the problem of unreliable synaptic vesicle release, which results in unreliable signal transmission between the pre-synaptic and the post-synaptic neurons (Rudolph et al. 2015). Several statistically independent unreliable contacts that sum together can reduce the variance of the post synaptic potentials (PSP). However, the same effect using a simpler mechanism could be achieved by multiple vesicles release (MVR) per synaptic activation (Rudolph et al. 2015; Holler et al. 2021) and does not require multiple synaptic contacts. Other studies addressed additional possible advantages for having multiple synaptic contacts between two neurons. Hiratani and Fukai (Hiratani and Fukai 2018) demonstrated that multiple synaptic contacts might allow synapses to learn quicker. Note that faster learning, although beneficial, fundamentally does not endow the neuron with the ability to perform new kinds of tasks. Zhang et al. (Zhang, Hu, and Liu 2020) model multiple contacts in the context of deep artificial neural networks but demonstrate no tangible computational benefit. Several other studies use multiple synaptic contacts in the context of artificial neural networks, demonstrating some computational

benefits, sometimes without explicitly addressing the use of multiple synaptic contacts (Camp, Mandivarapu, and Estrada 2020; Jones and Kording 2021; Sezener et al. 2021).

To address the question of the functional consequence of multiple contacts, we developed a simplified neuron model: the Filter and Fire (F&F) neuron. This model is based on the Integrate and Fire (I&F) neuron model but, in our model, each pre-synaptic axon makes multiple synaptic contacts. We further add features to the model to account for the effect of the dendritic cable filtering on time course of the somatic potential resulting from the different synaptic locations on the dendritic tree (Rall 1964; Rall 1967). To analyze the memory capacity of this model, we use the formulation of Memmesheimer et al. (Memmesheimer et al. 2014) developed for the I&F model. We further showed how to teach the F&F neuron a real-world classification task and explored other aspects such as the effect of unreliable synapses on the memory capacity of the F&F neuron and the implication of multiple synaptic contacts for optimizing axonal wiring in the brain.

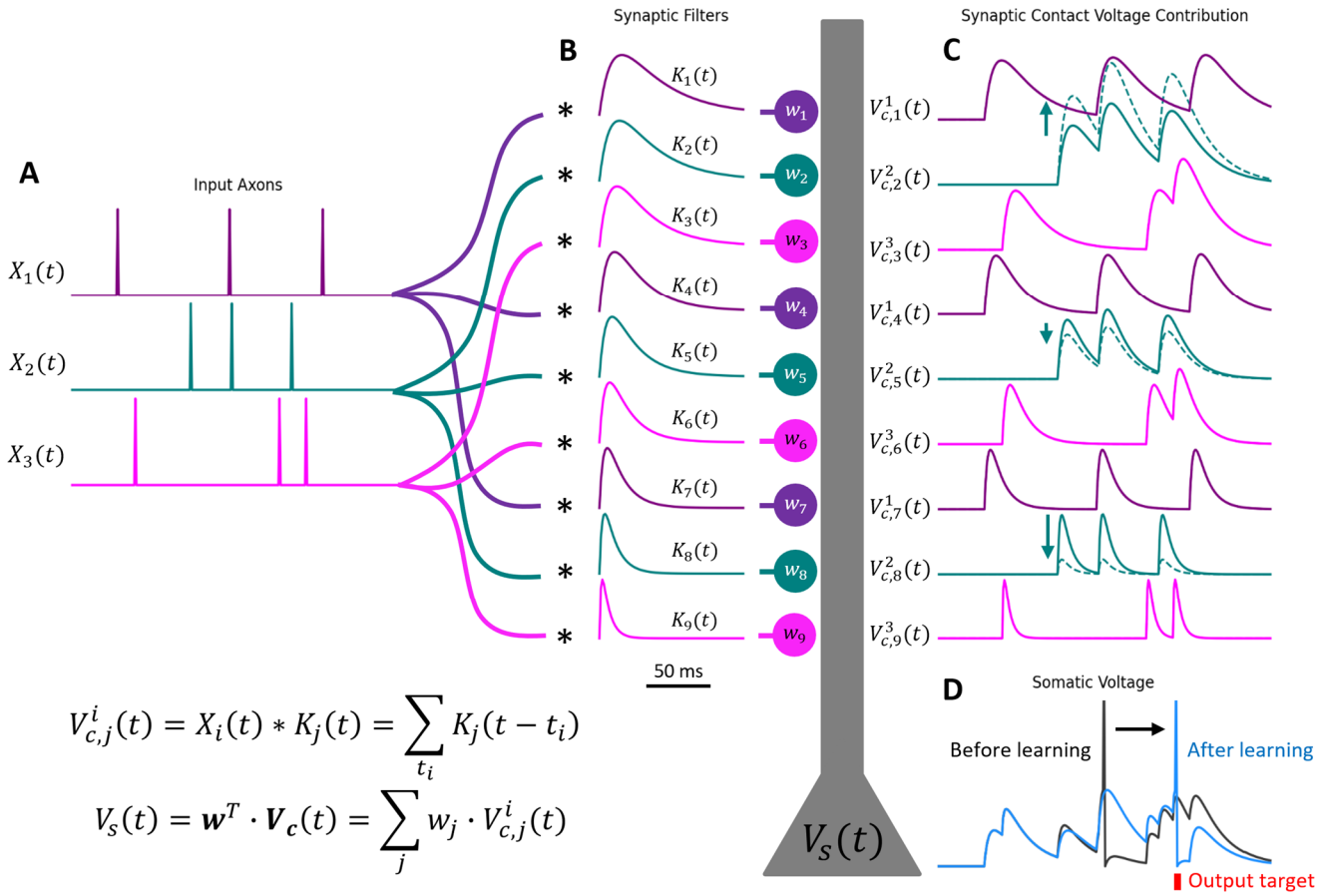
Results

Mathematical description of the filter and fire (F&F) neuron model with multiple contacts

We propose hereby a Filter and Fire (F&F) neuron model, which is similar to the standard current-based Leaky Integrate and Fire (I&F) neuron model, but with two added features. The first feature approximates the temporal characteristics of a dendritic cable as initially demonstrated by Rall (Rall 1964; Rall 1967), in which inputs that connect at distal locations on the dendrite exhibit prolonged post synaptic potentials (PSP) at the soma (Fig 1**B** top traces), whereas proximal inputs generates brief PSP profiles (Fig 1**B** bottom traces). The second feature is that each input axon connects to multiple locations on the dendritic tree, sometimes proximal and sometimes distal (Fig. 1**A**,**B)**. Formally, consider N_{axons} the number of input axons (Fig 1**A**), denoted by index i, and their spike trains will be represented by $X_i(t)$. Each axon connects to the dendrite via M contacts ($M = 3$ is illustrated in Fig. 1). Each contact connects to the dendrite at a location denoted by index j and filters the incoming axon spike train with a specific synaptic kernel $K_j(t)$. This forms the contact's voltage contribution trace $V_{c,j}^{i}(t) = X_i(t) * K_j(t) = \sum_{t_i} K_j(t - t_i)$. There is a total of $M \cdot N_{axons}$ such contact voltage contributions traces overall (Fig 1**C**). In vector notation we denote $\boldsymbol{V_c}(t) = [V_{c,1}^{1}(t), V_{c,2}^{2}(t), \cdots, V_{c,M\cdot N_{axons}}^{N_{axons}}(t)]$. Each synaptic contact has a weight, w_j. In vector notation we write $\boldsymbol{w} = [w_1, w_2, \cdots, w_{M\cdot N_{axons}}]$. Each contact contribution trace is multiplied by its corresponding weight to form the somatic voltage trace $V_s(t) = \boldsymbol{w}^T \cdot \boldsymbol{V_c}(t) =$

$\sum_j w_j \cdot V_{c,j}(t)$ (Fig 1**D**). When the spike threshold is reached, a standard reset mechanism is applied. Please note that the "dendrites" in this model are linear and therefore retain the analytic tractability of the I&F neuron models.

Here we model the temporal ramifications of the effect of adding a passive dendritic cable; we did not consider here the effect of nonlinear dendrites. The kernels we use are typical double exponential PSP shapes of the form: $K_j(t) = A \cdot \left(e^{-t/\tau_{decay,j}} - e^{-t/\tau_{rise,j}}\right)$, where A is a normalization constant such that each filter has a maximum value of 1, and $\tau_{decay,j}$, $\tau_{rise,j}$ are randomly sampled for each synaptic contact, representing randomly connected axon-dendrite locations. Note that due to mathematical simplicity we do not impose any restrictions on synaptic contact weights, each weight can be both positive or negative regardless of which axon it comes from. Indeed, the goal of the study is not to replicate all possible biological details, but specifically explore the computational benefit that arises due to two specific details - temporal filtering of synaptic potentials due to dendritic cable properties and multiple synaptic connections between pairs of neurons.



$$V^i_{c,j}(t) = X_i(t) * K_j(t) = \sum_{t_i} K_j(t - t_i)$$

$$V_s(t) = \boldsymbol{w}^T \cdot \boldsymbol{V_c}(t) = \sum_j w_j \cdot V^i_{c,j}(t)$$

Figure 1. The Filter and Fire (F&F) neuron receives input through multiple synaptic contacts per axon and filters each contact with a different synaptic kernel. (**A**) Example for three incoming input axons each making several contacts onto the post-synaptic cell. (**B**) Various synaptic filters representing their respective locations on the dendritic tree of the neuron model, shown in grey. Proximal synaptic filters are brief whereas more distal synaptic filters have broader temporal profiles. Colors are according to the source axon. (**C**) Local dendritic voltage responses at the synaptic loci, that result from the convolution of axonal spike train with the respective synaptic filter. Colors according to that of the source axon. Dashed lines indicate each contact's voltage contribution after learning. In this example, w_2 is increased, w_5 and w_8 are decreased following learning. (**D**) Somatic voltage is a weighted sum of each synaptic contact contributions with an independent weight for each contact. Standard I&F reset mechanism is used as the spike generation mechanism at the soma. In black is somatic trace before learning and in blue after learning.

Increased memory capacity of the F&F neuron with multiple synaptic contacts

We first test the memorization capacity of F&F neuron model as a function of the number of multiple connections. We utilize the framework proposed by Memmesheimer et al. (Memmesheimer et al. 2014) and measure memory capacity in an identical way, and use their proposed local perceptron learning rule for the task. In short, this capacity measure indicates the maximal number of precisely timed output spikes in response to random input stimulation during some time period. Fig. 2**A** shows random spiking

activity of 100 axons for a period of 60 seconds (top). Below the output of the post-synaptic cell is shown before learning (black), after learning (blue) and the desired target output spikes (red). In this example, we used $M = 5$ multiple contacts per axon. For the given set of spike trains, it is possible to find a weight vector to perfectly place all output spikes at their precisely desired timing. In Fig. 1**B** we repeat the simulation as in Fig. 1**A**, for various values of multiple contacts (M) while re-randomizing all input spike trains, desired output spike trains, and the synaptic filter parameters of each contact. We repeat this both for the I&F neuron model (i.e., a single synaptic PSP kernel for all synapses) and the F&F neuron model with a randomly selected synaptic kernel for each synapse (see **Methods** for full details of the kernel shapes we used). The y axis represents our success in placing all of the output spikes accurately, as measured by area under the receiver operating characteristic curve (ROC) (AUC) for the binary classification task of placing each spike in 1ms time bins. Error bars represent the variance of the AUC over multiple repeats (18), while re-randomizing the input, re-randomizing the synaptic kernels and re-randomizing the desired output spike trains. The figure shows that all the curves obtained from the I&F neuron models cluster together, and there is no change for different values of M (multiple contacts). This is to be expected as this is the classic case of synaptic redundancy when using a single temporal kernel for all pre-synaptic axons. e.g. for a single axon and two contacts one can see that in the case of a single synaptic kernel (as is the case in I&F model) the somatic voltage can be written as $V_s(t) = w_1 \cdot \sum K(t - t_i) + w_2 \cdot \sum K(t - t_i) = (w_1 + w_2) \cdot \sum K(t - t_i) = w_{eff} \cdot \sum K(t - t_i)$. i.e. the additional weights associated to the same input axon for the I&F neuron when using multiple contacts are equivalent to a single effective weight, and therefore not utilized.

For the F&F in the case of $M = 1$ we have an I&F model just with different kernels for each synapse. This change on its own does not make any difference in the capacity of the model as the number of learnable and utilizable parameters is identical to the I&F case with $M = 1$, and thus this curve lies with the other curves of the I&F models. However, for the F&F models with multiple contacts ($M = 2,3,5,10,15$) the graph shows an increased accuracy, demonstrating that some of the additional weights are utilized. In Fig. 2**C** we display the maximal number of output spikes that can be precisely timed as a function of M (number of multiple contacts), for both I&F and F&F models. We measure the number of precisely timed spikes as the maximal number of spikes that is above a high accuracy threshold (AUC > 0.99) of the plot in Fig. 2**B**, and we normalize by the number of axons to display the number of precisely timed output spikes per input axon on the y axis.

As expected from Fig. 1**B**, the capacity of the I&F model does not depend on the number of multiple connections. The F&F, however, displays an increase of capacity, saturating at approximately 3-fold larger than that of the I&F neuron model capacity. Note that the number of degrees of freedom (tunable parameters) scales linearly with the number of multiple contacts, so there is not an obvious explanation for the observed saturation. We will come back to this and explain the precise origin of this 3-fold increase compared to the I&F model in a later section. In Fig. 2D we vary the number of input axons and observe linear scaling of the number of precise spiking achieved with different slopes for different number of multiple connections. This indicates that increasing the number of multiple connections increases the effective number of parameters utilized per axon. To better illustrate these results, we show in Fig. S1**A** a simple case of how the I&F neuron can emit temporally precise output spikes by employing a spatial strategy. In Fig. S1**B** we show how an F&F neuron can employ a temporal strategy.

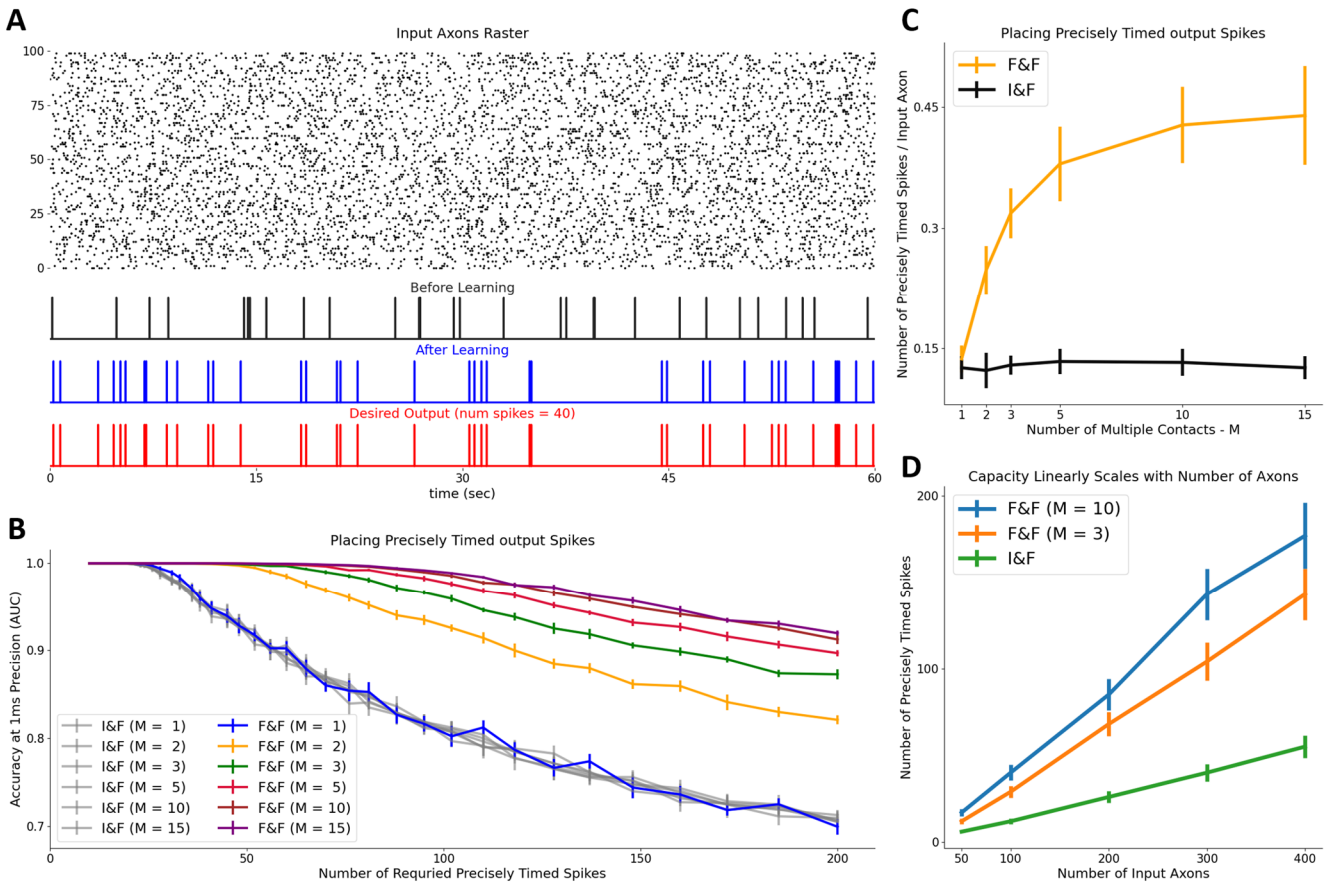


Figure 2. Increased memory capacity of precisely timed output spikes for F&F neuron compared to I&F neuron. (**A**) Learning to place precisely timed output spikes for randomly generated input. Top. Random axonal input raster. Bottom. Output spikes before learning (top, black), after learning (middle, blue) and the desired output spikes (bottom, red). (**B**) Binary classification accuracy at 1ms temporal resolution as measured by area under ROC curve (AUC) as a function of the number of required output spikes for input with 200 input axons. We see increased capacity for F&F models as the number of multiple contacts increases, whereas no increase for the I&F case as one would expect. (**C**) A summary plot that summarizes the capacity as a function of the number of multiple connections. For this plot we use the maximal number of spikes that achieves accuracy above AUC threshold of 0.99. The vertical axis depicts the fraction of successfully timed spikes for each input axon. We see saturation in the capacity for high values of multiple contacts to be ~3x compared to the I&F capacity. (**D**) The capacity scales linearly as function number of axons and exhibits no saturation.

The F&F neuron can learn spatio-temporal tasks that an I&F neuron cannot

Next, we wish to demonstrate new capabilities of the F&F neuron model with multiple synaptic connections that are beyond the I&F neuron model capabilities. For this purpose, we construct a new spatiotemporal task derived from MNIST task. Towards this end, we converted the horizontal spatial image dimension (width) into a temporal dimension (Fig. 3**A** top) with a uniform time warping such that 20 horizontal pixels will be mapped into T milliseconds. T will be the pattern presentation duration. The

vertical spatial image dimension (height) is simply replicated 5 times so that 20 vertical pixels will be mapped onto 100 axons. We then sample spikes for each axon according to the time varying Poisson instantaneous firing rate with additional background noise. An example of the resulting input spike trains is shown by the raster plot in Fig. 3**A** middle frame. Finally, we train our neuron to produce a spike at the end of a specific digit that was presented. We train on the full MNIST train subset of digits, and present results on the test set. Before learning (black), after learning (blue), and the desired output (red) are presented at the bottom of Fig. 3**A** for the case where the selected digit was 3. We then repeat this process for all digits for three models - I&F neuron, F&F neuron, and a spatio-temporal, temporally sliding, logistic regression (LR) model. We use the spatio-temporal sliding logistic regression model as a reference model for comparison with the F&F neuron but note that this model is not biologically plausible and cannot be considered as a model of a neuron in our case. Importantly, the LR model has $N_{axons} \cdot T$ learnable and fully utilizable parameters (N_{axons} parameters for each 1ms time bin) which is much greater than $N_{axons} \cdot M$ parameters that are used in the F&F and I&F models (which are also not fully utilizable as we've seen in Fig. 2).

The test accuracies following training for all models are depicted in Fig 3**B**. For this plot a successful true positive (hit) is achieved if at least 1 spike has occurred in the time window of 10 ms around the ground truth desired spike. The temporal duration of each pattern T was 40 ms and the number of multiple contacts M was 5, for the plot in Fig. 3**B**. Fig. 3**B** clearly shows that the I&F neuron model is at chance level for almost all digits, and basically is incapable of learning the task. In contrast, the F&F model with 5 multiple connections is consistently better than chance and sometimes approaches the "aspirational" spatio-temporal logistic regression model. In Fig. 3**C** we visually display the learned weights of all models when attempting to learn the digit 3. The weight matrix of the logistic regression model clearly depicts what appears to be an average-looking digit 3, which was the digit the neuron was trained to recognize. The F&F neuron model depicts a temporally smoothed version of the logistic regression model, and the I&F model clearly cannot learn temporal patterns and therefore cannot recognize this digit at above chance level. For precise details of how the weights for the F&F and I&F models were visualized, please consult the **Methods**. For a more simplified pattern classification case, please see Fig. S1**C**. Fig. S1**D** shows how a F&F neuron can solve the task shown in Fig. S1**C**. Fig. S1**E** explains why this task cannot be solved by an I&F neuron.

To test the effect of the presentation duration of each digit T, Fig. 3**D** displays summary statistics of test accuracy, averaged across all digits, for the three models as a function of T. The interval between patterns was 70 ms and the decay time constant for I&F model was 30 ms, to match the maximal decay time

constant for all synaptic kernels in the F&F model. The F&F neuron had 5 multiple contacts. Fig. 3**D** shows that there is an optimal pattern presentation duration for the F&F model that occurs at the 40-50 ms range, which is ~1.5 times the maximal decay time constant in the model.

Next, we seek to determine the effect of unreliable synaptic transmission and its interaction with multiple synaptic contacts. As explained in the Introduction, multiple synaptic contacts are often considered as a mechanism to overcome synaptic transmission unreliability. In Fig. 3**E** we display the test accuracy as a function of the number of multiple contacts for all 3 models for fully reliable synapses as we have displayed thus far, and in dashed line is the accuracy under the unreliable synapse's regime with a probability of release of $p = 0.5$ for each contact. Note the consistent drop in test accuracy that does not go away even with a large number of multiple contacts. In this graph the positive digit used was "7" and the pattern temporal duration was $T = 30$ ms. The number of positive training samples used in this case was 2048 patterns. Unreliable synaptic transmission can be considered as a mechanism for implementing "drop-connect", a method used in training artificial neural networks that has a known regularization effect (Wan et al. 2013; Srivastava et al. 2014). In Fig. 3**F** we test if this is also applicable to our case. We show the test accuracy as a function of the number of training samples and demonstrate a regularization effect in the case of F&F that increases test accuracy for a low number of training input patterns. Note that each training sample was shown to the model multiple times (15 times in this case) and for each spike and each contact an independent probability of release was applied, effectively resulting in 15 noisy patterns that were presented to the neuron during training for each original training pattern. This suggests that unreliable synapses can also be viewed as a "feature" rather than a "bug" for the regime of a small number of training data points and can help avoid overfitting.

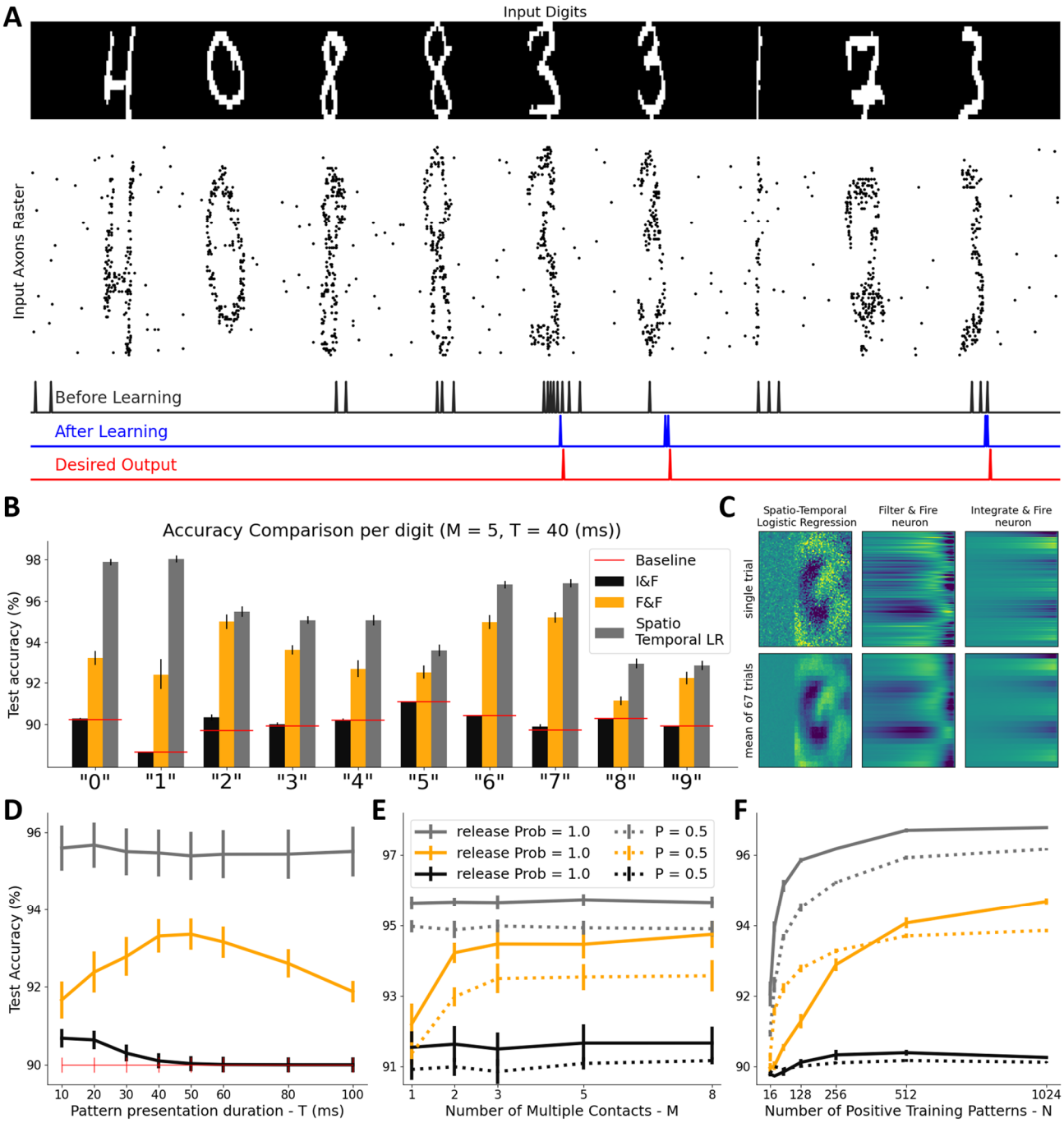


Figure 3. The F&F neuron can learn to recognize spatio-temporal patterns whereas I&F neuron cannot. (**A**) A F&F neuron was trained to recognize digits in a spatio-temporal version of the MNIST task. From top to bottom: Original input digits, axon spike raster that represent the respective digits (with some additional background noise), F&F neuron (with 5 multiple contacts) output before learning (black), after learning, (blue) and the desired output spike train (red). The neuron was trained to detect the digit "3" in this case. The depicted traces are from unseen test set. (**B**) Full test set classification accuracy of F&F (orange), I&F (black) and a non-biologically plausible temporally sliding spatio-temporal Logistic Regression model (gray) for the spatio-temporal digit classification task when each digit was used as the positive class and the rest of the digits were considered the negative class. The random chance baseline for each case is shown by the red horizontal lines. The I&F models cannot learn the task whereas F&F neurons sometimes approach the (non-biological) spatio-temporal logistic

regression model. The number of positive training samples used in this case was 5000. (**C**) Spatio-temporal representation of the learned weights of the three models when attempting to detect the digit "3". We can clearly discern the digit "3" in the logistic regression case, a faint "3" in the F&F case, and an attempt to use spatial-only information to detect the digit in the I&F case, which is unsuccessful in achieving above chance accuracy on the test set. The top three images are for a single trial with a single random axonal wiring, and Poisson sampling of the inputs. Bottom three images are an average of 67 trials. (**D**) Summary of test accuracy, averaged across all digits, for the three models as a function of the temporal duration of the pattern. Silence time between patterns was 70ms. The decay time constant for I&F model is 30 ms, as is the maximal decay time constant for all synaptic kernels for the F&F model. F&F neuron in this case had 5 multiple contacts. (**E**) Test accuracy as a function of the number of multiple contacts for all 3 models. In dashed line is the accuracy under the unreliable synapse's regime with a probability of release of 50% for each contact. Note the consistent drop in accuracy. In this graph the positive digit used was "7" and the pattern temporal duration was T=30ms. The number of positive training samples used in this case was 2048 patterns. (**F**) Test accuracy as a function of the number of training samples. Each training sample was shown to the model multiple times (15 times in this case) and for each spike and each contact an independent probability of release was applied. Note the regularization benefit of unreliable synaptic release for a small number of training samples in the dataset (dashed orange line above solid orange line). This is simple result of drop-connect regularization that unreliable synapses implement. This suggests that unreliable synapses can be viewed as a "feature" rather than a "bug" for the regime of a small number of training data points.

Maximal capacity is explained by the effective 3-dimensional subspace spanning all synaptic kernels

Next, we wish to pinpoint the mathematical origin of the properties depicted in Figures 2 and 3 of the proposed F&F neuron models with multiple synaptic connections. We observed that all PSP kernels we used are of similar shapes and therefore the inputs from the same axon will generate very correlated inputs at the local synaptic responses vector $\boldsymbol{V_c}(t)$. Any input correlation will limit the number of degrees of freedom available for learning by modification of the synaptic weights $\boldsymbol{w}$. In Fig. 4**A** we show all the PSPs as heatmaps organized according to increasing values of τ_{rise} within each block and increasing τ_{decay} between the blocks. In Fig. 4**C** we show all the PSPs as temporal traces overlayed on each other. Both Fig. 4**A** and 4**C** clearly show that the shapes of the PSP kernels, although chosen randomly and have some variance, are overall very similar to each other. We therefore apply singular matrix (SVD) decomposition on all PSP shapes (Fig. 4**B**) and discover that 99.93% of the variance in all PSP shapes is explained by the first 3 singular vectors (Fig. 4**E**). This means all synaptic kernels depicted in Fig. 4A and 4C are effectively spanned by a basis set of orthogonal 3 PSP-like shapes. For the sake of presentation and to avoid negative values in the trace shapes, we display the 3 independent kernels that are the result of non-negative matrix factorization (NMF) in Fig. 4**D**. It's visible that these PSP shapes basically filter the input signal with various time constants and various delays. These are very intuitive shapes that we

can easily interpret. With them, we can understand both the temporal smoothing aspect of the learned weights by the F&F neuron model in Fig. 3**C**, and the number of independent PSP shapes is a good candidate to explain the very specific 3-fold increase in capacity of Fig. 2**C**. To verify that these the basis sets are in fact explaining the phenomenon in Fig. 2, we repeat the same experimentation as in Fig. 2, but now each axon is connected to the post synaptic neuron via only 3 multiple connections, and this time they are the optimal PSPs shapes depicted in Fig. 4**D** instead of being randomly selected. These results are depicted in Fig. 4**F**. The model with three orthogonal kernels has effectively identical results to those of the F&F neuron, and thus helps us explain its behavior. First, to explain the capacity, when we randomly sample more and more connections from the set of all possible kernels (e.g., increasing M), we slowly approach to span the entire 3-dimensional space by each contact and therefore we have a saturation effect at three times the I&F baseline level. This means that we have reached the three independent kernels, each corresponding to an independent learnable parameter. Second, the specific shapes of these kernels explain the temporal smoothing effect we observe in the learned weights of the F&F neuron model in Fig. 3**C**. Note that although we used in our study somewhat artificial double exponential synaptic kernels, we verified that our results also hold for PSPs in a simulation of a highly realistic detailed cortical neuron model. This verification can be found in Fig. S2, which shows that despite some quantitative differences, qualitatively the results presented in our work using a reduced F&F neuron model are valid also for the case of a neuron with a complex dendritic tree. Namely, a basis set of three temporally-distinct kernels can effectively span all synaptic kernels in realistic models of cortical pyramidal neurons.

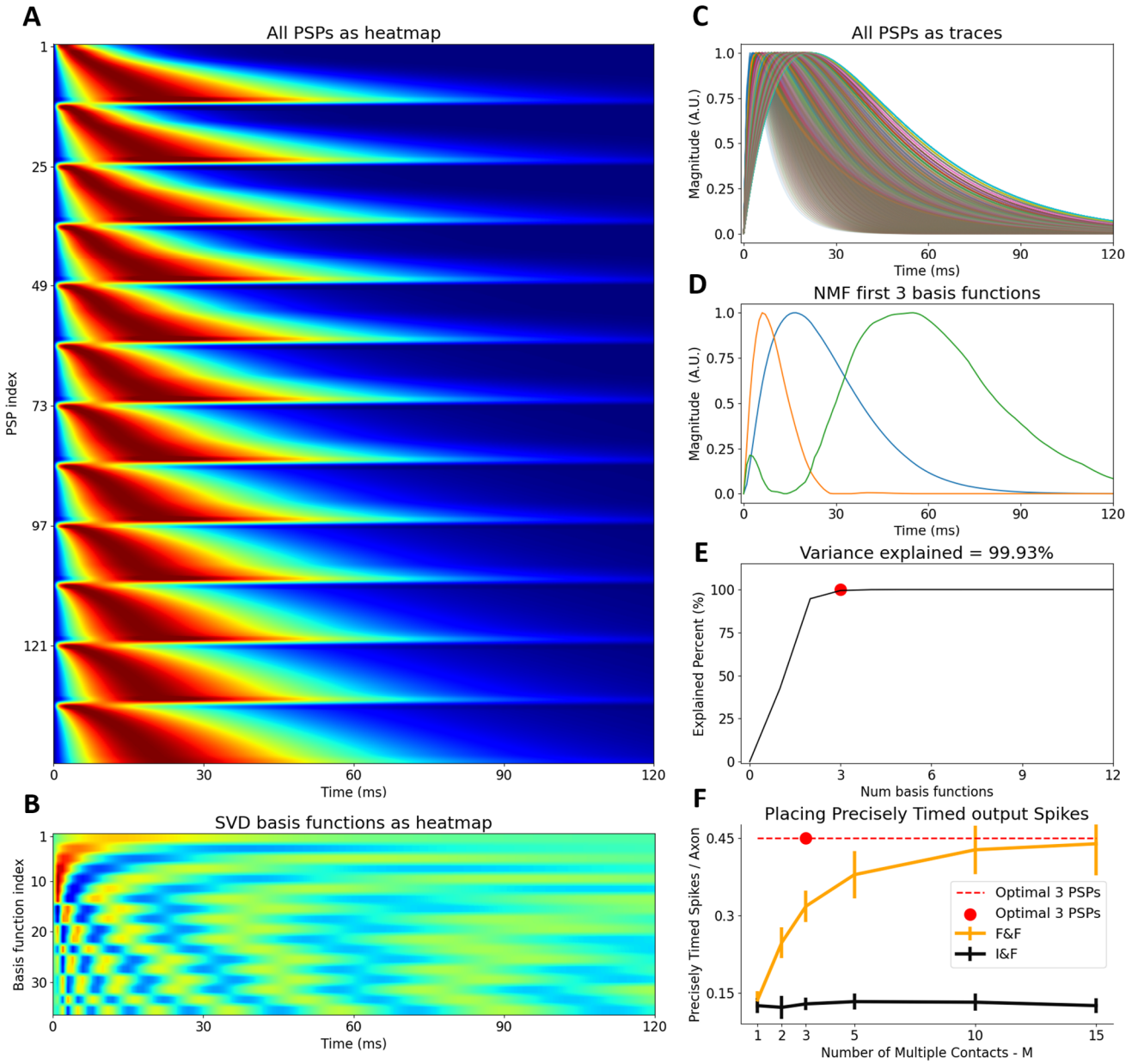


Figure 4. The dendritic filters are spanned by a 3-dimensional basis set of PSPs accounting for the 3-fold increase in the F&F capacity for a large number multiple contacts. (**A**) All possible post synaptic potentials (PSPs) that were used in the study shown as heatmaps. Every block has a different exponential decay time. Within each block the rise time grows for each row**.** (**B**) The Singular Value Decomposition of all the PSPs shown in (A) as heatmaps. We see a Fourier-like basis set. (**C**) All possible post synaptic potentials (PSPs) that were used in the study, shown as traces. We clearly see that there are strong correlations between the various PSPs, indicating they are not all equally useful (**D**) First 3 Basis functions of the Non-Negative Matrix Decomposition of all PSPs shown in (A) and (C). NMF instead of SVD was used for ease of interpretation and visualization purposes. (**E**) Cumulative variance explained by each basis component. We see that by using 3 basis functions shown in (D) we can span all PSPs shown in (C) and explain the phenomenon shown in Fig. 2**C** in which the capacity is capped by 3x the number of axons when the number of multiple contacts is large. (**F**) Direct verification that the 3 orthogonal basis set in (E) can be used as 3 optimal multiple contact filters and achieve the same capacity as 15 randomly selected multiple contacts.

Multiple synaptic contacts along with dendrites allow the CNS to save either in number of transmitting axons or decoding neurons while achieving identical computational goals

Finally, we wish to point out that multiple synaptic contacts and the proposed F&F neuron model might allow the central nervous system to transmit information via axons in a more hardware efficient way by relying on dendritic decoding downstream. Three computationally equivalent alternatives are depicted in Figure 5. Fig. 5**C** illustrates the case in which N axons transmit information to be readout by an F&F neuron model with multiple synaptic contacts downstream. If we wish for a downstream I&F point neuron to have the ability to produce the same output as the F&F in Fig. 5**C**, we will need that each of the N axons in Fig. 5**C** to be replicated 3 times, each with some additional delay to account for the temporal integration properties of the F&F neuron model. This scenario is illustrated in Fig. 5**D** and verified in simulations in Figures 5**A** and 5**B**. Alternatively, it is also possible to utilize the same N axons as in Fig. 5**C** but have a decoding network of 3N point neurons downstream. This scenario is depicted in Fig. 5**E**. These three alternatives highlight the more general scenario in which the evolutionary pressure to reduce axonal wiring in an ever-increasing brain volume (Chklovskii, Schikorski, and Stevens 2002; Chen, Hall, and Chklovskii 2006) is solved by compressing information on a limited number of axons and relying on sophisticated dendritic integration to decode these signals. In this context we have shown that employing multiple synaptic contacts between the axon and its post-synaptic neuron enables dendrites to better decode spatiotemporal patterns and save "hardware" (reduced total axon length, reduced number of neurons).

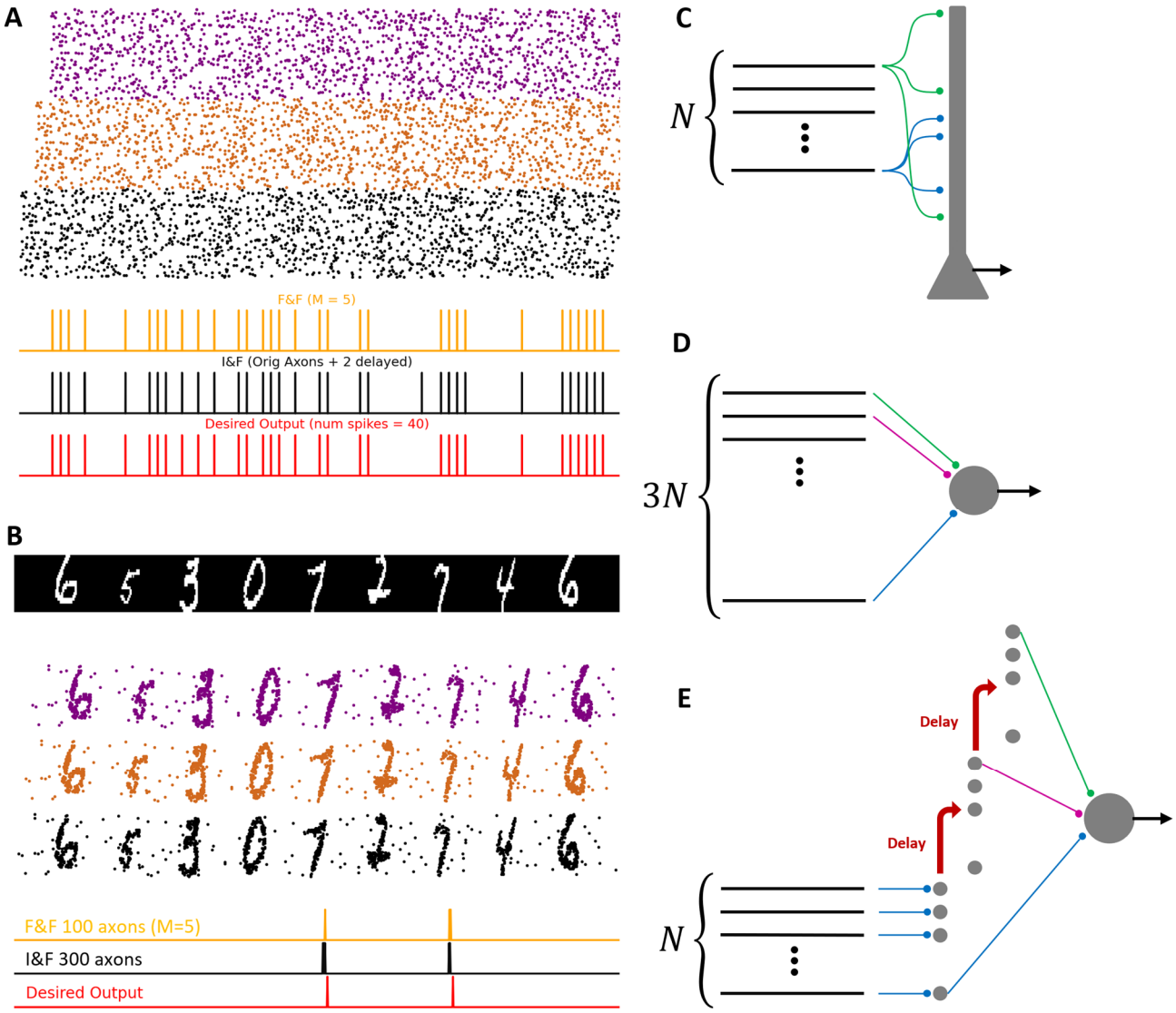


Figure 5. The F&F neuron model allows the central nervous system to reduce the number of transmitting axons or decoding neurons. Illustration of three computationally equivalent implementations. (**A**) On the top depicted is the input axon raster. The black axons are delayed once to produce the brown axons and delayed once more to produce the purple axons. Only the black axons are fed as input to the F&F neuron, whose output is shown in orange. The full set of axons is fed into a I&F neuron whose output is shown in black. In red the desired output spike train is shown. This panel demonstrates that an I&F neuron with additionally delayed input axons replicates the precise timing capacity of the F&F model. (**B**) Similar to what is depicted in (A) only this time for the spatiotemporal MNIST digit recognition task. Here again we see that an I&F neuron with additionally delayed input axons can reproduce the F&F neuron performance on the spatiotemporal MNIST digit recognition task (**C**) Illustration of N axons that feed into a single F&F neuron model with a simplified dendrite and multiple synaptic contacts. (**D**) In order for an I&F point neuron model to have identical memorization capacity as the F&F neuron illustrated in (C), the same information needs to be separated into 3N axons (**E**) In order for an I&F point neuron to have identical spatio-temporal patterns separation capabilities with the same original N axons as in (C), it needs to have a feed-forward "delay line" decoding network prior to being fed into the I&F point neuron. This alternative is demonstrated in both (A) and (B).

Discussion

Neurons in the brain typically connect to each other via multiple synaptic contacts. The computational role of this redundancy is unclear. In this work we augmented the commonly used Leaky Integrate and Fire (I&F) neuron model with simplified dendrites that filter the incoming spike train via a set of multiple post synaptic potential (PSP) filters. Each filter corresponds to a particular synaptic contact, with varied time constants that directly relate to signal filtering by the cable properties of the dendritic tree. This is in contrast to a single PSP filter for all contacts in the I&F point neuron model. We term this new neuron model Filter and Fire (F&F). Specifically, we have demonstrated that the capacity of the F&F neuron to memorize precise input-output relationships is increased by a factor of ~3 compared to that of the regular I&F. The capacity is measured as the ratio between the number of precisely timed output spikes and the number of incoming input axons. This ratio is ~0.15 for the I&F case as shown by Memmesheimer et al. (Memmesheimer et al. 2014); and grew to ~0.45 spikes per axons for the F&F model.

Next, we constructed a new spatiotemporal pattern discrimination task using the MNIST dataset and demonstrated that our neuron model can learn to detect single digits at well-above chance level performance on an unseen test set, whereas an I&F neuron model cannot learn the task at all, because in the specific way we chose to represent each digit – the task does not contain enough spatial-only information suitable for I&F neuron discrimination. Our specific task design was deliberately chosen to highlight this temporal aspect of pattern discrimination that is possible when using taking into account the temporal filtering due to cable properties of dendrites. We show that multiple synaptic connections with different PSP profiles allow the neuron to effectively parametrize the temporal profile of the PSP influence of each pre-synaptic axon on the somatic membrane potential. This is enabled by modifying the weight of the various (multiple) contacts made between the axon and the post synaptic cell. We show that all PSPs can be spanned by a 3 basis PSP filters, each with a different temporal profile. Taken together, this suggests that the F&F neuron model provides a low temporal frequency approximation to a spatio-temporal perceptron that assigns independent weights to each point in time. An alternative description that is mathematically equivalent, is that the F&F model effectively bins the membrane integration time into 3 non-uniform time bins and can learn to assign independent weights for each bin in the past.

Our study demonstrates that even when considering highly simplified neuron model as used here, that implements the passive temporal filtering aspect of dendrites, a computational role of dendrites is unraveled for the seeming redundancy of multiple synaptic contacts between pairs of neurons in the brain.

Dendrites therefore allow us to salvage some of the redundant connectivity and put those synaptic weight parameters to good use. This allows for increased memorization capacity, but perhaps more importantly to detect specific spatiotemporal patterns in the input axons. The spatiotemporal filtering properties of dendritic processing are prominently featured in a recently published study describing how single neurons produce output spikes in response to highly complex spatiotemporal patterns (Beniaguev, Segev, and London 2021). It is important to note that there is an additional nonlinear amplification aspect in dendrites (Poirazi and Mel 2001; Poirazi, Brannon, and Mel 2003; Polsky, Mel, and Schiller 2004; Bicknell and Häusser 2021) that we did not consider in this study and that is likely to provide additional computational benefits also in the context of multiple synaptic contacts, as the previous work by Beniaguev et al. greatly suggests (Beniaguev, Segev, and London 2021). We have therefore added in the present study a new perspective on the growing literature regarding the computational function of individual neuron (McCulloch and Pitts 1943; Rosenblatt and F. 1958; Oja 1982; Hyvarinen and Oja 1998; Moldwin and Segev 2018; Golkar et al. 2020; Pehlevan et al. 2020; Gütig and Sompolinsky 2006a; Poirazi, Brannon, and Mel 2003; Häusser and Mel 2003; Moldwin, Kalmenson, and Segev 2021; Zador, Claiborne, and Brown 1991; Mel 1992; Poirazi and Mel 2001)

It is notable that it is typically believed that the phenomenon of multiple synaptic contacts is primarily attributed to noise reduction related to unreliable synaptic transmission, but our work strongly suggests that this is only a small part of the story from a computational standpoint. In fact, we would like to propose that each synapse can reduce its own noise by employing a multiple vesicle release (MVR) tactic if needed, as was also recently suggested by (Rudolph et al. 2015). If this is in fact the case, it is possible that unreliable synaptic transmission might play a computational role of its own and rather than being an unwanted "bug", it might turn out to be a useful "feature". More specifically, it might play a role that is similar to dropout (Srivastava et al. 2014) or more precisely drop-connect (Wan et al. 2013) which are commonly used in present day artificial neural networks paradigm. There, drop-connect is typically employed as a regularization technique that reduces overfitting and usually improves generalization. Although this was not the main focus of our work, we briefly demonstrate this effect in Figure 3**F**.

The question of how trains of spikes represent information in the nervous system has been a long-standing question in neuroscience, since its inception. A major debate revolves around whether information is largely carried by firing rates averaged over relatively long time periods, or rather that precisely timed spikes carry crucial bits of information. Evidence for both alternatives has been found for both sensory systems and motor systems and much theoretical work on this key topic has been conducted (Meister,

Lagnado, and Baylor 1995; Christopher Decharms and Merzenich 1996; Wehr and Laurent 1996; Neuenschwander and Singer 1996; Johansson and Birznieks 2004; Hopfield 1995; Castelo-Branco et al. 2000; Thorpe, Delorme, and Van Rullen 2001; DeWeese, Wehr, and Zador 2003; Kara, Reinagel, and Reid 2000; London et al. 2010; Abeles et al. 1993; Schneidman, Freedman, and Segev 1998; Memmesheimer et al. 2014; Florian 2012; London et al. 2002; Abeles 1982; Gütig and Sompolinsky 2006b; Maass and Schmitt 1999). Since dendrites and multiple synaptic connections play such a crucial role in decoding incoming spike trains and increase the neuronal repertoire in emitting precisely times output spikes in response to spatiotemporal input patterns (and they might do this using a simple biologically plausible learning rule), we wish to suggest that dendritic hardware should be considered when discussing the question of the neural code and what information is transmitted via axons. In (Perez-Nieves et al. 2021), the authors show that a diversity of time constants helps increasing the computational repertoire of spiking networks.

Here we suggest that a similar thing might happen at the neuronal level and not only at the network level. The fact that a single neuron can decode complex spatiotemporal patterns on its own and does not require a highly coordinated decoding network of neurons to extract temporal information from incoming spike trains, not only allows for potential "hardware" savings as we illustrated in Fig. 5, but also suggest that information might be ubiquitously encoded by precise spike times throughout the central nervous system. A single neuron can emit precisely timed output spikes in response to spatiotemporal inputs, as we showed here and was previously shown by (Memmesheimer et al. 2014). It is therefore not required to have a large and highly coordinated network of neurons to encode temporally precise patterns that can be sent via axons. The fact that a single neuron can do this on it's own without the use of network mechanisms, somewhat increases the likelihood of information being encoded by precise spike timing as opposed to average firing rates (over relatively long periods of time) throughout the CNS.

Lastly, in recent years, multiple groups around the world have started to create detailed reconstructions of rodent and human brains, and report neuronal connectivity maps with the help of electron microscopy (EM) (Kasthuri et al. 2015; Motta et al. 2019; Shapson-Coe et al. 2021). We would like to suggest that analyzing these EM datasets, focusing on the number of multiple contacts and their locations on the dendritic tree, might shed some additional light on the extent and role of multiple synaptic contacts between different cell types and brain regions, and hint to the possible “style” of information processing in the network based solely on EM data. As illustrated by our work, if two neurons form a connection on distal dendrites, or if they form a connection on proximal dendrites, these will result in completely different influences (time course) of the somatic voltage and, therefore, on the temporal coding

capabilities of the neuron. Indeed, simply reporting connectivity maps and even the size of post synaptic density (PSD) areas is not enough to determine the temporal influence of the connection on the post synaptic cell, as the dendritic location of the synapse is key, as was also suggested in (Liu et al. 2021). Also, due to the nonlinear amplification of dendrites, it will be crucial to know whether two pre-synaptic neurons connect to similar locations on the dendritic tree as they are much more likely to undergo nonlinear amplification (and might produce additional broad/slow NMDA or Calcium dependent temporal filters) if activated at similar times.

Methods

F&F neuron simulation details

A F&F neuron receives as input N_{axons} input axons, their spike trains will be represented by $X_i(t) = \sum_{t_i} \delta(t - t_i)$ and they will be denoted by index i. Each axon connects to the dendrite via M contacts Each contact connects on the dendrite on a location that will be denoted by index j, and filters the incoming axon spike train by with a specific synaptic kernel $K_j(t)$. The kernels are typical double exponential PSP shapes of the form $K_j(t) = A \cdot \left(e^{-t/\tau_{decay,j}} - e^{-t/\tau_{rise,j}}\right)$ where A is a normalization constant such that each filter has a maximum value of 1, and $\tau_{decay,j}$, $\tau_{rise,j}$ are different for each contact, sampled randomly and independently for each contact from the ranges $\tau_{decay,j} \in [12ms, 30ms]$, $\tau_{rise,j} \in [1ms, 12ms]$. Different kernel parameters represent a randomly connected axon-dendrite location.

The result of the kernel filtering of the corresponding input axons forms the contact voltage contribution trace $V_{c,j}^{i}(t) = X_i(t) * K_j(t) = \sum_{t_i} K(t - t_i)$. There are a total of $M \cdot N_{axons}$ such contact voltage contributions traces overall. In vector notation we denote $\boldsymbol{V_c}(t) = [V_{c,1}(t), V_{c,2}(t), \cdots, V_{c,M \cdot N_{axons}}(t)]$. Each synaptic contact has a weight, w_j. In vector notation we write $\boldsymbol{w} = [w_1, w_2, \cdots, w_{M \cdot N_{axons}}]$. each local synaptic response is multiplied by its corresponding weight to form the somatic voltage $V_s(t) = \boldsymbol{w}^T \cdot \boldsymbol{V_c}(t) = \sum_j w_j \cdot V_{c,j}(t)$. When threshold is reached, the voltage is reset, and a negative rectifying current is injected that decays to zero with time constant of 15ms. Note that due to mathematical simplicity we do not impose any restrictions on synaptic contact weights, each weight can be both positive or negative regardless of which axon it comes from.

I&F neuron simulation details

The I&F simulation details are identical in all ways to the F&F neuron, except that all of it's contact kernels are identical $K_j(t) = K_{I\&F}(t) = A_{I\&F} \cdot \left(e^{-t/\tau_{decay,I\&F}} - e^{-t/\tau_{rise,I\&F}}\right)$, where $\tau_{rise,I\&F} = 1ms$ and $\tau_{decay,I\&F} = 30ms$.

Capacity calculation experimental details

To measure capacity, we sample N_{axons} random spike trains to serve as axons from Poisson instantaneous firing rate of 4Hz for a period of 120 seconds. We randomly distribute N_{spikes} output spikes throughout the 120 second time period to generate $y_{GT}(t)$. For the sake of mathematical simplicity and not dealing with reset issues, we make sure that the minimal distance between two consecutive spikes is at least 120ms ($= 4 \cdot \tau_{decay,I\&F}$). We bin time to 1ms time bins. We calculate $\boldsymbol{V_c}(t)$ for the entire trace. Our task is to find a set of weights such that $y_{GT}(t) = \varphi(\boldsymbol{w}^T \cdot \boldsymbol{V_c}(t))$, where $\varphi(\cdot)$ is a simple thresholding function. This is in essence a binary classification dataset with 120,000 timepoints (milliseconds), for each of those timepoints the required output of a binary classifier is either 1 (for time points that should emit an output spike), or 0 for all other timepoints. We have a total $M \cdot N_{axons}$ weights that need to fit the entire 120K samples dataset. We calculate the AUC on the entire 120K datapoints dataset. We declare that the fit was successful when AUC > 0.99. We repeat the procedure for various values of N_{spikes}. The maximal value of N_{spikes} that we still manage to fit with AUC > 0.99 is termed $N_{spikes,max}$ and the capacity measure is this number normalized by the number of axons used $Capacity(M) = N_{spikes,max}/N_{axons}$. Note that due to $N_{spikes} \ll 120{,}000$, the capacity of the problem effectively doesn't depend on the length of the time period we use, but by rather the number of spikes we wish to precisely time.

Spatio-temporal MNIST task details

The MNIST dataset contains 60,000 training images and 10,000 test images. The images are of size 28x28 pixels. We crop the images at the center to be 20x20 pixels, and binarize the values. We the conver the horizontal spatial image dimension (width) into a temporal dimension by uniformly warping the time such that 20 horizontal pixels will be mapped into T milliseconds. T is the pattern presentation duration. The vertical spatial image dimension (height) is simply replicated 5 times so that 20 vertical pixels will be mapped into 100 axons. The entire training set is the concatenated sequentially (in random ordering), with 70ms of zeros between every two patterns. We then sample spikes for every 1ms time bin according to each axon's instantaneous firing rate to generate the raster for the entire training set. On top of that we

add an additional background noise rate. The output ground truth is a single output spike 1ms after the pattern is presented for the positive class digit, and no spikes for negative patterns. The fitting of the model is identical to that described in the capacity sections. i.e., we wish to find a set of weights such that $y_{GT}(t) = \varphi(\boldsymbol{w}^T \cdot \boldsymbol{V_c}(t))$, where $\varphi(\cdot)$ is a simple thresholding function. In this case, we allow for wiggle room when evaluating performance on the test set. A successful true positive (hit) is achieved if at least 1 spike has occurred in the time window of 10ms around the ground truth desired spike. A failed false positive (false alarm) is considered if a spike has occurred during the time window of 10ms around the end of the pattern presentation. We measure the classification accuracy under these criteria. We sometimes train only on a part of the training dataset and use only $N_{positive\ samples}$ as the positive class. In the regime of unreliable synaptic transmission with some synaptic release probability p, we sample the spikes of the input patterns once, and then present it $N_{epochs} = 15$ times, each time with random release probability samples for each pre-synaptic spike for each contact. In those cases, we perform the same procedure also for the test set (i.e., display the same pattern $N_{epochs} = 15$ times, each time with different synaptic release sampled for each pre-synaptic spike for each contact)

Spatio-temporal Logistic Regression (LR) details

Spatio-temporal temporally sliding logistic regression (LR) model is a non-biologically plausible that is specifically used to serve as an aspirational model used for comparison for our proposed F&F neuron. It has a 2-dimensional weight matrix $W_{LR}(s,t)$ where s spans the spatial dimention and t spans the temporal dimension. Mathematically, the equation describing the relationship between the input and the output of the model is given by $y_{LR}(t) = \varphi(\sum_{s=0}^{N_{axons}} \sum_{\tau=0}^{T_{LR}} W_{LR}(s,\tau) \cdot X_s(s, t-\tau))$. Note that this model has in total $N_{axons} \cdot T_{LR}$ weights (time is discretized into 1ms time bins here as well, as throughout this study).

F&F learned weights visualization

In the F&F model a single input axon is filtered by multiple contact kernels $V_{c,j}^{i}(t) = X_i(t) * K_j(t)$. In order to display the effective linear model weights, we can group together all kernels that relate to the same axon $V^i(t) = \sum_{j_i} w_{j_i} \cdot X_i(t) * K_{j_i}(t) = X_i(t) * \sum_{j_i} w_{j_i} \cdot K_{j_i}(t)$. Therefore, the term $\sum_{j_i} w_{j_i} \cdot K_{j_i}(t)$ is a composite kernel that filters each axon. This is a function that can be visualized for each input axons, and therefore related directly to the input space.

Code and data availability

All code necessary to reproduce all results of this paper are available on GitHub via the link: https://github.com/SelfishGene/filter_and_fire_neuron

All data and live scripts to reproduce all figures are available on Kaggle at the following link: https://www.kaggle.com/selfishgene/fiter-and-fire-paper

Acknowledgements

We thank all lab members of the Segev and London Labs for many fruitful discussions and valuable feedback. This work was supported by the Drahi family foundation [https://plfa.info/], a grant from the ETH domain for the Blue Brain Project [https://www.epfl.ch/research/domains/bluebrain/], the Gatsby Charitable Foundation [gatsby.org.uk] and the NIH Grant Agreement U01MH114812 (I.S.).

Supplementary Figures

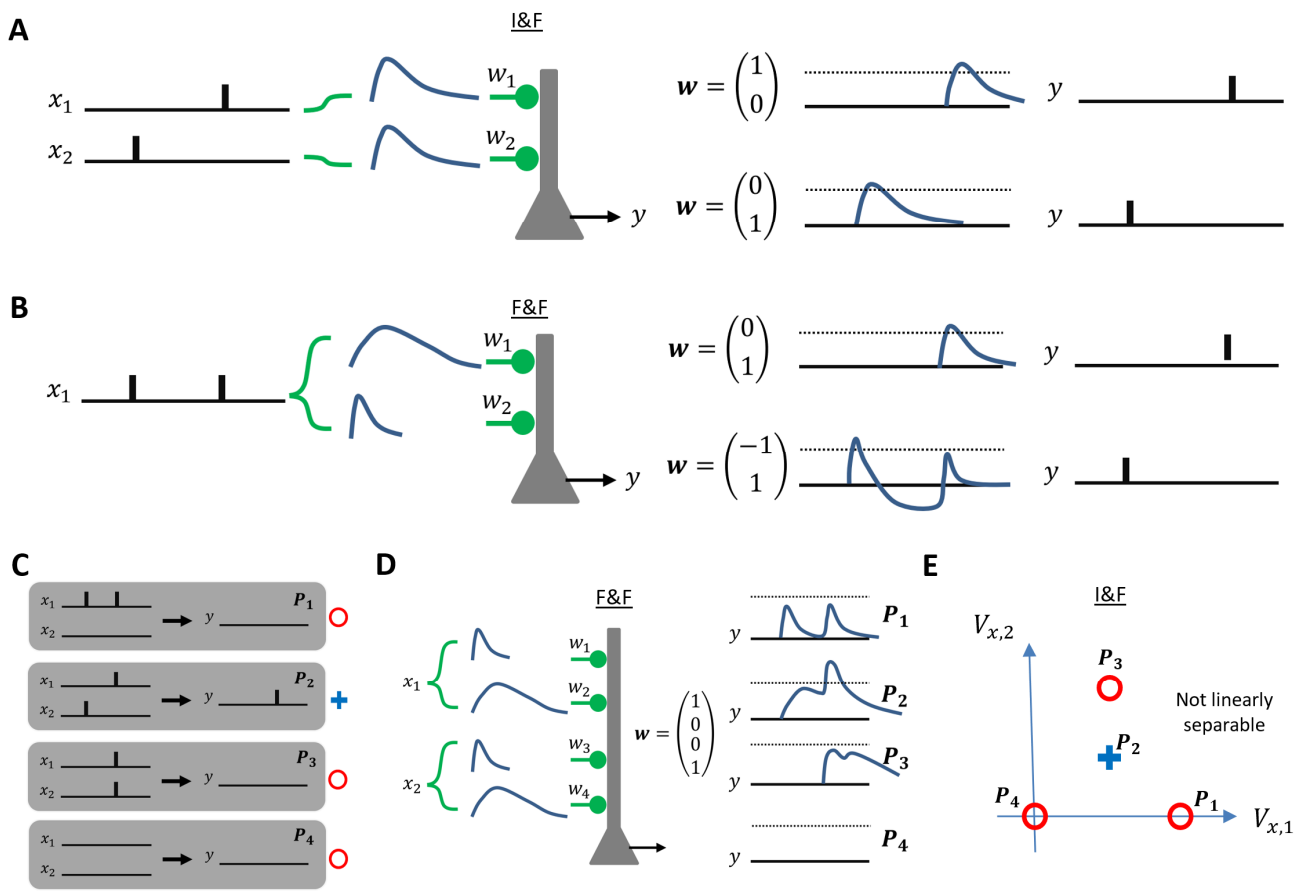


Figure S1. Toy illustration to highlight differences in computational capabilities between the F&F neuron and I&F neuron. (**A**) Illustration of the spatial strategy the I&F model employs. Left. Input axons x1, x2 are fed into a I&F model to emit an output y. Right, two output for different weight configurations that result in two different output patterns y. Dashed line represent spike threshold (**B**) Illustration of the temporal strategy the F&F model employs. Left. A single input axon x1 is fed into an F&F neuron with two different PSP filters to produce output y. Right, two output for different weight configurations that result in two different output patterns y. note that although the output patterns are identical to those in (A), this was achieved by receiving a single input axon. (**C**) Example of a temporal task we wish to teach the two neuron models. For the four different input patterns (x1, x2) we require to produce the output y on the right. (patterns are denoted as P1, P2, P3, P4). Only P2 is required to emit an output spike (marked with blue plus sign), the other patterns are required to emit no spikes (marked with red circle) (**D**) The solution of a F&F model with fast and slow PSP filters for each input axon. On the right are the somatic responses for each pattern with the weights w. Dashed line represent spike threshold. (**E**) The I&F neuron model cannot classify the output patterns as the four patterns are not linearly separable in the dendritic representation space (Vx,1 Vx,2) as seen in the illustration.

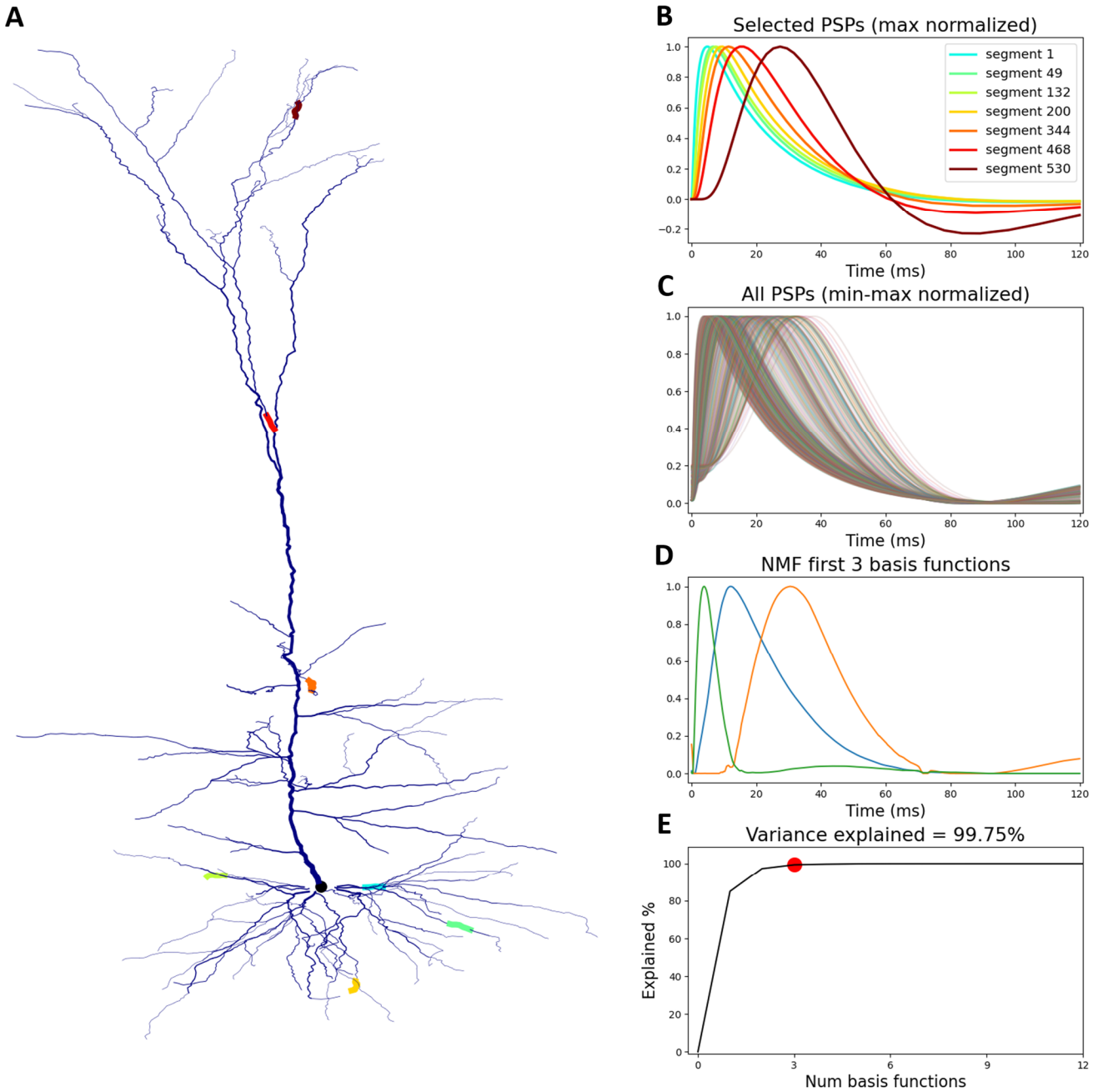


Figure S2. The dendritic filters of a reconstructed L5PC detailed biophysical model are also spanned by a 3-dimensional basis set of PSPs as does our simplified F&F neuron throughout the paper. (**A**) The morphology of a L5PC that was biophysically modeled in detailed (Hay et al. 2011) with highlighted dendritic segments**.** (**B**) Normalized somatic PSPs in response to excitatory synaptic input at the highlighted dendritic segments in (A). Note that in some of the EPSPs there exists a small hyperpolarization that is due to nonlinear potassium channels at the soma (**C**) Normalized somatic EPSPs from all dendritic segments in the modeled L5PC. The traces are normalized such that minimal value in the time window is 0 and maximal value is 1 (**D**) First 3 Basis functions of the Non-Negative Matrix Decomposition of all EPSPs shown in (C). NMF instead of SVD was used for simplicity of interpretation and visualization purposes. These filters look very similar to those shown in Fig. 4**D**. (**E**) Cumulative variance explained by each basis component demonstrating that the 3 basis functions shown in (D) can span all EPSPs shown in (C).

References

Abeles, M. 1982. "Role of the Cortical Neuron: Integrator or Coincidence Detector?" *Israel Journal of Medical Sciences* 18 (1): 83–92.

Abeles, M., H. Bergman, E. Margalit, and E. Vaadia. 1993. "Spatiotemporal Firing Patterns in the Frontal Cortex of Behaving Monkeys." *Https://Doi.Org/10.1152/Jn.1993.70.4.1629* 70 (4). American Physiological Society Bethesda, MD : 1629–38. doi:10.1152/JN.1993.70.4.1629.

Acharya, Jyotibdha, Arindam Basu, Robert Legenstein, Thomas Limbacher, Panayiota Poirazi, and Xundong Wu. 2021. "Dendritic Computing: Branching Deeper into Machine Learning." *Neuroscience*, October. Pergamon. doi:10.1016/J.NEUROSCIENCE.2021.10.001.

Beniaguev, David, Idan Segev, and Michael London. 2021. "Single Cortical Neurons as Deep Artificial Neural Networks." *Neuron*, August. Cell Press. doi:10.1016/J.NEURON.2021.07.002.

Bicknell, Brendan A., and Michael Häusser. 2021. "A Synaptic Learning Rule for Exploiting Nonlinear Dendritic Computation." *Neuron* 109 (24). Cell Press: 4001–4017.e10. doi:10.1016/J.NEURON.2021.09.044.

Camp, Blake, Jaya Krishna Mandivarapu, and Rolando Estrada. 2020. "Continual Learning with Deep Artificial Neurons," November. https://arxiv.org/abs/2011.07035v1.

Castelo-Branco, Miguel, Rainer Goebel, Sergio Neuenschwander, and Wolf Singer. 2000. "Neural Synchrony Correlates with Surface Segregation Rules." *Nature 2000 405:6787* 405 (6787). Nature Publishing Group: 685–89. doi:10.1038/35015079.

Chen, Beth L, David H Hall, and Dmitri B Chklovskii. 2006. "Wiring Optimization Can Relate Neuronal Structure and Function." www.wormatlas.org.

Chklovskii, Dmitri B., Thomas Schikorski, and Charles F. Stevens. 2002. "Wiring Optimization in Cortical Circuits." *Neuron* 34 (3). Cell Press: 341–47. doi:10.1016/S0896-6273(02)00679-7.

Christopher Decharms, R., and Michael M. Merzenich. 1996. "Primary Cortical Representation of Sounds by the Coordination of Action-Potential Timing." *Nature* 381 (6583): 610–13. doi:10.1038/381610A0.

DeWeese, Michael R., Michael Wehr, and Anthony M. Zador. 2003. "Binary Spiking in Auditory Cortex." *Journal of Neuroscience* 23 (21). Society for Neuroscience: 7940–49. doi:10.1523/JNEUROSCI.23-21-07940.2003.

Fares, Tarec, and Armen Stepanyants. 2009. "Cooperative Synapse Formation in the Neocortex." *Proceedings of the National Academy of Sciences* 106 (38). National Academy of Sciences: 16463–68. doi:10.1073/PNAS.0813265106.

Feldmeyer, Dirk, Joachim Lübke, and Bert Sakmann. 2006. "Efficacy and Connectivity of Intracolumnar Pairs of Layer 2/3 Pyramidal Cells in the Barrel Cortex of Juvenile Rats." *Journal of Physiology* 575 (2). J Physiol: 583–602. doi:10.1113/jphysiol.2006.105106.

Florian, Răzvan V. 2012. "The Chronotron: A Neuron That Learns to Fire Temporally Precise Spike Patterns." *PLOS ONE* 7 (8). Public Library of Science: e40233. doi:10.1371/JOURNAL.PONE.0040233.

Golkar, Siavash, David Lipshutz, Yanis Bahroun, Anirvan M Sengupta, and Dmitri B Chklovskii. 2020. "A Simple Normative Network Approximates Local Non-Hebbian Learning in the Cortex."

Gütig, Robert, and Haim Sompolinsky. 2006a. "The Tempotron: A Neuron That Learns Spike Timing–based Decisions." *Nature Neuroscience 2006 9:3* 9 (3). Nature Publishing Group: 420–28. doi:10.1038/nn1643.

———. 2006b. "The Tempotron: A Neuron That Learns Spike Timing-Based Decisions." *Nature Neuroscience* 9 (3). Nature Publishing Group: 420–28. doi:10.1038/nn1643.

Häusser, Michael, and Bartlett Mel. 2003. "Dendrites: Bug or Feature?" *Current Opinion in Neurobiology* 13 (3). Elsevier Current Trends: 372–83. doi:10.1016/S0959-4388(03)00075-8.

Hay, Etay, Sean Hill, Felix Schürmann, Henry Markram, and Idan Segev. 2011. "Models of Neocortical Layer 5b Pyramidal Cells Capturing a Wide Range of Dendritic and Perisomatic Active Properties." Edited by Lyle J. Graham. *PLoS Computational Biology* 7 (7): e1002107. doi:10.1371/journal.pcbi.1002107.

Hiratani, Naoki, and Tomoki Fukai. 2018. "Redundancy in Synaptic Connections Enables Neurons to Learn Optimally." *Proceedings of the National Academy of Sciences* 115 (29). National Academy of Sciences: E6871–79. doi:10.1073/PNAS.1803274115.

Holler, Simone, German Köstinger, Kevan A.C. Martin, Gregor F.P. Schuhknecht, and Ken J. Stratford. 2021. "Structure and Function of a Neocortical Synapse." *Nature* 591 (7848). Nature: 111–16. doi:10.1038/s41586-020-03134-2.

Hopfield, J. J. 1995. "Pattern Recognition Computation Using Action Potential Timing for Stimulus Representation." *Nature 1995 376:6535* 376 (6535). Nature Publishing Group: 33–36. doi:10.1038/376033a0.

Hyvarinen, Aapo, and Erkki Oja. 1998. "Independent Component Analysis by General Nonlinear Hebbian-like Learning Rules." *Signal Processing* 64: 301–13. doi:10.1016/s0165-1684(97)00197-7.

Johansson, Roland S, and Ingvars Birznieks. 2004. "First Spikes in Ensembles of Human Tactile Afferents Code Complex Spatial Fingertip Events." *Nature Neuroscience 2004 7:2* 7 (2). Nature Publishing Group: 170–77. doi:10.1038/nn1177.

Jones, Ilenna Simone, and Konrad Paul Kording. 2021. "Might a Single Neuron Solve Interesting Machine Learning Problems Through Successive Computations on Its Dendritic Tree?" *Neural Computation* 33 (6). MIT Press: 1554–71. doi:10.1162/NECO_A_01390.

Kara, Prakash, Pamela Reinagel, and R. Clay Reid. 2000. "Low Response Variability in Simultaneously Recorded Retinal, Thalamic, and Cortical Neurons." *Neuron* 27 (3). Elsevier: 635–46. doi:10.1016/S0896-6273(00)00072-6.

Kasthuri, Narayanan, Kenneth Jeffrey Hayworth, Daniel Raimund Berger, Richard Lee Schalek, José Angel Conchello, Seymour Knowles-Barley, Dongil Lee, et al. 2015. "Saturated Reconstruction of a Volume of Neocortex." *Cell* 162 (3): 648–61. doi:10.1016/j.cell.2015.06.054.

Liu, Tony X., Pasha A. Davoudian, Kristyn M. Lizbinski, and James M. Jeanne. 2021. "Connectomic Features Underlying Diverse Synaptic Connection Strengths and Subcellular Computation." *BioRxiv*, August. Cold Spring Harbor Laboratory, 2021.08.19.456845. doi:10.1101/2021.08.19.456845.

London, Michael, Arnd Roth, Lisa Beeren, Michael Häusser, and Peter E. Latham. 2010. "Sensitivity to Perturbations in Vivo Implies High Noise and Suggests Rate Coding in Cortex." *Nature 2010 466:7302* 466 (7302). Nature Publishing Group: 123–27. doi:10.1038/nature09086.

London, Michael, Adi Schreibman, Michael Haä;usser, Matthew E. Larkum, and Idan Segev. 2002. "The Information Efficacy of a Synapse." *Nature Neuroscience 2002 5:4* 5 (4). Nature Publishing Group: 332–40. doi:10.1038/nn826.

Maass, Wolfgang, and Michael Schmitt. 1999. "On the Complexity of Learning for Spiking Neurons with Temporal Coding." *Information and Computation* 153 (1). Academic Press: 26–46. doi:10.1006/INCO.1999.2806.

Markram, H, J Lübke, M Frotscher, A Roth, and B Sakmann. 1997. "Physiology and Anatomy of Synaptic Connections between Thick Tufted Pyramidal Neurones in the Developing Rat Neocortex." *The Journal of Physiology* 500 (Pt 2). Wiley-Blackwell: 409. doi:10.1113/JPHYSIOL.1997.SP022031.

Markram, Henry, Eilif Muller, Srikanth Ramaswamy, Michael W Reimann, Marwan Abdellah, Carlos Aguado Sanchez, Anastasia Ailamaki, et al. 2015. "Reconstruction and Simulation of Neocortical Microcircuitry." *Cell* 163 (2). Elsevier: 456–92. doi:10.1016/j.cell.2015.09.029.

McCulloch, Warren S., and Walter Pitts. 1943. "A Logical Calculus of the Ideas Immanent in Nervous

Activity." *The Bulletin of Mathematical Biophysics* 5 (4). Kluwer Academic Publishers: 115–33. doi:10.1007/BF02478259.

Meister, Markus, Leon Lagnado, and Denis A. Baylor. 1995. "Concerted Signaling by Retinal Ganglion Cells." *Science* 270 (5239). Science: 1207–10. doi:10.1126/science.270.5239.1207.

Mel, Bartlett W. 1992. "NMDA-Based Pattern Discrimination in a Modeled Cortical Neuron." *Neural Computation* 4: 502–17. doi:10.1162/neco.1992.4.4.502.

Memmesheimer, Raoul-Martin, Ran Rubin, Bence P Olveczky, and Haim Sompolinsky. 2014. "Learning Precisely Timed Spikes." *Neuron* 82 (4). Neuron: 925–38. doi:10.1016/J.NEURON.2014.03.026.

Moldwin, Toviah, Menachem Kalmenson, and Idan Segev. 2021. "The Gradient Clusteron: A Model Neuron That Learns to Solve Classification Tasks via Dendritic Nonlinearities, Structural Plasticity, and Gradient Descent." *PLOS Computational Biology* 17 (5). Public Library of Science: e1009015. doi:10.1371/JOURNAL.PCBI.1009015.

Moldwin, Toviah, and Idan Segev. 2018. "Perceptron Learning and Classification in a Modeled Cortical Pyramidal Cell." *BioRxiv*, 464826. doi:10.1101/464826.

Motta, Alessandro, Manuel Berning, Kevin M. Boergens, Benedikt Staffler, Marcel Beining, Sahil Loomba, Philipp Hennig, Heiko Wissler, and Moritz Helmstaedter. 2019. "Dense Connectomic Reconstruction in Layer 4 of the Somatosensory Cortex." *Science* 366 (6469). American Association for the Advancement of Science. doi:10.1126/SCIENCE.AAY3134.

Neuenschwander, Sergio, and Wolf Singer. 1996. "Long-Range Synchronization of Oscillatory Light Responses in the Cat Retina and Lateral Geniculate Nucleus." *Nature 1996 379:6567* 379 (6567). Nature Publishing Group: 728–33. doi:10.1038/379728a0.

Oja, E. 1982. "A Simplified Neuron Model as a Principal Component Analyzer." *Journal of Mathematical Biology* 15 (3): 267–73. doi:10.1007/BF00275687.

Pehlevan, Cengiz, Xinyuan Zhao, Anirvan M. Sengupta, and Dmitri Chklovskii. 2020. "Neurons as Canonical Correlation Analyzers." *Frontiers in Computational Neuroscience* 14 (June). Frontiers Media S.A.: 55. doi:10.3389/FNCOM.2020.00055/BIBTEX.

Perez-Nieves, Nicolas, Vincent C. H. Leung, Pier Luigi Dragotti, and Dan F. M. Goodman. 2021. "Neural Heterogeneity Promotes Robust Learning." *BioRxiv*, March. Cold Spring Harbor Laboratory, 2020.12.18.423468. doi:10.1101/2020.12.18.423468.

Peters, Alan, and Martin L. Feldman. 1976. "The Projection of the Lateral Geniculate Nucleus to Area 17 of the Rat Cerebral Cortex. I. General Description." *Journal of Neurocytology* 5 (1). J Neurocytol: 63–84. doi:10.1007/BF01176183.

Poirazi, Panayiota, Terrence Brannon, and Bartlett W Mel. 2003. "Pyramidal Neuron as Two-Layer Neural Network." *Neuron* 37 (6). Elsevier: 989–99. doi:10.1016/S0896-6273(03)00149-1.

Poirazi, Panayiota, and Bartlett W. Mel. 2001. "Impact of Active Dendrites and Structural Plasticity on the Memory Capacity of Neural Tissue." *Neuron* 29 (3). Cell Press: 779–96. doi:10.1016/S0896-6273(01)00252-5.

Polsky, Alon, Bartlett W Mel, and Jackie Schiller. 2004. "Computational Subunits in Thin Dendrites of Pyramidal Cells." *Nature Neuroscience* 7 (6): 621–27. doi:10.1038/nn1253.

Rall, W. 1967. "Distinguishing Theoretical Synaptic Potentials Computed for Different Soma-Dendritic Distributions of Synaptic Input." *Journal of Neurophysiology* 30 (5): 1138–68. doi:10.1152/jn.1967.30.5.1138.

Rall, Wilfrid. 1964. "Theoretical Significance of Dendritic Trees for Neuronal Input-Output Relations." *Neural Theory and Modeling*. Stanford University Press, 73–97. https://ci.nii.ac.jp/naid/10025255881/.

Rees, Christopher L., Keivan Moradi, and Giorgio A. Ascoli. 2017. "Weighing the Evidence in Peters' Rule: Does Neuronal Morphology Predict Connectivity?" *Trends in Neurosciences* 40 (2). NIH Public Access: 63. doi:10.1016/J.TINS.2016.11.007.

Rosenblatt, F., and F. 1958. “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain.” *Psychological Review* 65 (6). American Psychological Association: 386–408. doi:10.1037/h0042519.

Rudolph, Stephanie, Ming-Chi Tsai, Henrique von Gersdorff, and Jacques I. Wadiche. 2015. “The Ubiquitous Nature of Multivesicular Release.” *Trends in Neurosciences* 38 (7). NIH Public Access: 428. doi:10.1016/J.TINS.2015.05.008.

Schneidman, Elad, Barry Freedman, and Idan Segev. 1998. “Ion Channel Stochasticity May Be Critical in Determining the Reliability and Precision of Spike Timing.” *Neural Computation*. Vol. 10. http://www.weizmann.ac.il/neurobiology/labs/schneidman/Publications/Schneidman+al_1998-NeuralComp.pdf.

Sezener, Eren, Agnieszka Grabska-Barwińska, Dimitar Kostadinov, Maxime Beau, Sanjukta Krishnagopal, David Budden, Marcus Hutter, et al. 2021. “A Rapid and Efficient Learning Rule for Biological Neural Circuits.” *BioRxiv*, March. Cold Spring Harbor Laboratory, 2021.03.10.434756. doi:10.1101/2021.03.10.434756.

Shapson-Coe, Alexander, Michał Januszewski, Daniel R. Berger, Art Pope, Yuelong Wu, Tim Blakely, Richard L. Schalek, et al. 2021. “A Connectomic Study of a Petascale Fragment of Human Cerebral Cortex.” *BioRxiv*, June. Cold Spring Harbor Laboratory, 2021.05.29.446289. doi:10.1101/2021.05.29.446289.

Shepherd, Gordon M G, Armen Stepanyants, Ingrid Bureau, Dmitri Chklovskii, and Karel Svoboda. 2005. “Geometric and Functional Organization of Cortical Circuits.” *Nature Neuroscience 2005 8:6* 8 (6). Nature Publishing Group: 782–90. doi:10.1038/nn1447.

Silver, R. Angus, Joachim Lübke, Bert Sakmann, and Dirk Feldmeyer. 2003. “High-Probability Uniquantal Transmission at Excitatory Synapses in Barrel Cortex.” *Science* 302 (5652). Science: 1981–84. doi:10.1126/science.1087160.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting.” *Journal of Machine Learning Research* 15 (56): 1929–58. http://jmlr.org/papers/v15/srivastava14a.html.

Thorpe, Simon, Arnaud Delorme, and Rufin Van Rullen. 2001. “Spike-Based Strategies for Rapid Processing.” *Neural Networks* 14 (6–7). Neural Netw: 715–25. doi:10.1016/S0893-6080(01)00083-1.

Wan, Li, Matthew Zeiler, Sixin Zhang, Yann Le Cun, and Rob Fergus. 2013. “Regularization of Neural Networks Using DropConnect.” PMLR. https://proceedings.mlr.press/v28/wan13.html.

Wehr, Michael, and Gilles Laurent. 1996. “Odour Encoding by Temporal Sequences of Firing in Oscillating Neural Assemblies.” *Nature 1996 384:6605* 384 (6605). Nature Publishing Group: 162–66. doi:10.1038/384162a0.

Zador, Anthony M, Brenda J Claiborne, and Thomas H Brown. 1991. “Nonlinear Pattern Separation in Single Hippocampal Neurons with Active Dendritic Membrane.” In *NIPS*.

Zhang, Jiangshe, Junying Hu, and Junmin Liu. 2020. “Neural Network with Multiple Connection Weights.” *Pattern Recognition* 107 (November). Pergamon: 107481. doi:10.1016/J.PATCOG.2020.107481.