bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

An efficient analytical reduction of nonlinear detailed neuron models

[1]Oren Amsalem✉*, [1]Guy Eyal*, [1]Noa Rogozinski* and [1, 2]Idan Segev

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

* These authors contributed equally to this work
✉ Correspondence: Oren Amsalem - oren.amsalem1@mail.huji.ac.il

Keywords: Neuron_Reduce, Compartmental modeling, Nonlinear dendrites, Cortical synapses, Single neuron computation, Cable theory

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Detailed conductance-based neuron models, consisting of nonlinear branched dendrites and thousands of synapses, are essential for understanding the integrative and computational properties of single neurons and large neuronal networks, and for interpreting experimental results. Simulations of such models are computationally expensive, severely limiting their utility. We introduce a novel analytic approach to simplify complex nonlinear neuron models while preserving the identity of individual dendrites and synapses. *Neuron_Reduce* represents each stem dendrite by a unique cylindrical cable, keeping its specific membrane and axial properties. *Neuron_Reduce* maps synapses and active membrane ion channels to the respective cylinder while preserving their transfer impedance to- and from- the soma as in the detailed model. The reduced model accelerates the simulation speed by up to 200-fold while closely replicating the sub- and supra- threshold voltage dynamics for a variety of cell types and inputs, including the nonlinear "ping pong" interaction between somatic Na^+- and dendritic Ca^{2+}-spikes, found in L5 neocortical pyramidal cells. *Neuron_Reduce* also replicates dendritic computations discriminating spatiotemporal input sequences. The reduced neuron models will enable realistic simulations of neural networks at unprecedented scale, including of biologically-inspired "deep networks" and facilitate the construction of neuromorphic-based systems. *Neuron_Reduce* is publicly available (https://github.com/orena1/neuron_reduce) and is straightforward to implement.

Introduction

Compartmental models (CMs) were first employed to study neurons by Wilfrid Rall (Rall, 1964). They enabled him to explore the impact of spatio-temporal activation of conductance-based dendritic synapses on the neuron's output, and the effect of the dendritic location of a synapse on the shape (time course) of its excitatory post-synaptic potentials at the soma (Rall, 1967). By simulating realistic (electrically-distributed) neuron models, Rall demonstrated how the cable properties of dendrites can explain the variety of somatic EPSPs shapes that were recorded at the soma of α-motoneurons, negating, the dominant explanation at that time, that the different shape of the somatic EPSPs in these cells was attributed to differences in the kinetics of the respective synapses. This was an impressive example showing that a faithful model of the neuron (as a distributed rather than a "point" electrical unit) is essential for the correct interpretation of experimental results. Since Rall's 1964 and 1967 studies using CMs, the EPSP "shape indices" has become a standard method for estimating the location of dendritic synapses from experimentally measured somatic EPSPs.

Over the years, compartmental models have provided key insights into hundreds of experimental findings, both for the single cell level and the network level. A notable example at the single neuron level is the explanation as to why the somatic Na^+ action potential propagates backward from soma-to-dendrites and (typically) not *vice versa* (Rapp et al., 1996); CMs also pinpointed the conditions for the generation of dendritic

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ca^{2+} spike (Hay et al., 2011; Larkum et al., 2009) and provided the explanation for the spatial restriction of active spread of dendritic spikes from distal dendrites to the soma (Segev, 1992) and see also (Bahl et al., 2012; Gouwens et al., 2018; Magee and Cook, 2000; Migliore et al., 1999; Poirazi et al., 2003; Segev and London, 1999; Stuart and Spruston, 1998). At the network level, noteworthy examples are the use of detailed CMs by (Egger et al., 2014; Markram et al., 2015) for large-scale simulations of a densely *in silico* reconstructed cortical circuits. Another team at the Allen Institute recently started a 10-year mission for understanding signal flow/computations in the mouse visual system and are using detailed CMs for simulating large parts of the visual system of the mouse (Hawrylycz et al., 2016). Indeed, because detailed compartmental modeling is increasingly becoming an essential tool for understanding diverse neuronal phenomena, major efforts have been invested in developing user-friendly computer software that implements detailed CMs for the use of the neuroscientific community, most notable are NEURON (Carnevale and Hines, 2006), GENESIS (Bower, 1998), NeuroConstruct (Gleeson et al., 2007), Geppetto (Cantarelli et al., 2018) and NTS (Kozloski and Wagner, 2011).

Present-day personal computers enable the simulations of tens of seconds of the activity of single neurons each comprising of thousands of nonlinear compartments with realistic numbers (many thousands) of synapses. However, computer simulations of neural networks consisting of many thousands of such neurons require computational power that is beyond what is available today at most laboratories. When the size of the network reaches hundreds of thousands of such neurons, very powerful computers are required to run the simulation. For example, in the Blue Brain Project, simulation of a cortical network consisting of 200,000 neurons (Markram et al., 2015) requires several hours using a BlueGene/Q machine for simulating 30 sec of real time.

To overcome this limitation, two approaches are typically employed. One is to develop alternative, cheaper and more efficient computing architectures (e.g., neuromorphic-based computers (Aamir et al., 2016; Schemmel et al., 2008)). These have not yet ripened to simulate large-scale network models with neurons consisting of branched nonlinear dendrites. The other approach is to simplify the neuron models while preserving, as faithfully as possible, their input/output relationship. Rall (Rall, 1962) was the first to provide such a scheme in his “equivalent cylinder” model where he showed that, for certain idealized passive dendritic trees, the whole tree could be

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

collapsed to a single cylinder, that is analytically identical to the full tree. The "equivalent cylinder" preserves the total dendritic membrane area, the electrotonic length of the full dendritic tree and the input resistance at the soma (Rall and Rinzel, 1973; Rinzel and Rall, 1974). This method is not applicable for dendritic trees with large variability in the dendritic cable lengths (e.g., pyramidal neurons with long apical tree and short basal tree) or for dendritic trees with nonlinear membrane properties.

Along the years, several additional reduction schemes have been proposed; e.g., a recent work (Rössert et al., 2016) mapped all synapses to a single compartment, taking into account the filtering effect of the dendrites. Other methods reduced the full morphology to a simplified geometrical model while preserving the total membrane area (Destexhe, 2001; Hendrickson et al., 2011; Stratford et al., 1989) or the axial resistivity (Bush and Sejnowski, 1993); see also (Bahl et al., 2012; Marasco et al., 2012). However, these methods are either "hand fit" without a clear analytical underpinning or are complicated to implement, and their computational speedup for realistic numbers of synapses was not demonstrated. Most of them do not support dendrites with active conductances (Brown et al., 2011; Bush and Sejnowski, 1993; Marasco et al., 2012, 2013; Stratford et al., 1989). Importantly, none of the previous studies supplied an easy-to-use open access implementation for their respective methods, and they were not tested on a variety of neuron types. We, therefore, lack a simple, publicly-available, reduction method for neuron models to be used by the large neuroscience and machine-learning communities.

The present work provides an analytic method for reducing the complexity of detailed neuron models while faithfully preserving the essential input/output properties of these models. *Neuron_Reduce* is based on key insights from Rall's cable theory, and its implementation for any neuron type is straightforward without the need for hand fitting. Depending on the neuron modeled and the number of synapses, *Neuron_Reduce* accelerates the run-time by a factor of up to 200 while preserving the identity of individual synapses and their respective dendrites. It also preserves the specific membrane properties and dendritic nonlinearities, thus preserving specific dendritic computations. *Neuron_Reduce* is easy to use, is fully documented and is publicly available at GitHub (https://github.com/orena1/neuron_reduce).

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Analytic mapping of the full dendritic tree to a reduced multi-cylinder tree

The thrust of our analytical reduction method (*Neuron_Reduce*) is described in **Figure 1A-C**. This method is based on reducing each of the original dendrites, stemming from the soma, to a single cylindrical cable having the same specific membrane resistivity (R_m, in Ωcm^2), capacitance (C_m, in F/cm^2), and axial resistivity (R_i, in Ωcm), as that of the full tree (**Figure 1A**). In addition, the cylindrical cable satisfies two constraints: (i) the magnitude of transfer impedance, $|Z_{0,L}(\omega)| = |V_0(\omega)/I_L(\omega)|$, from its distal sealed end ($X = L$) to its origin ($X = 0$) is identical to the magnitude of the transfer impedance from the electrotonically most distal dendritic tip to the soma in the respective original dendrite; (ii) at the proximal end of that cylinder (X = 0), the magnitude of the input impedance, $|Z_{0,0}(\omega)| = |V_0(\omega)/I_0(\omega)|$, matches that of the respective stem dendrite (when decoupled from the soma). These two constraints, while preserving the specific membrane and axial properties, ensure a unique cylindrical cable (with a specific diameter and length) for each of the original dendrites (see Eqs. (1) – (11) in **Methods**)

As the transfer impedance in both the original dendrite and in the respective cylindrical cable spans from $|Z_{0,L}(\omega)|$ to $|Z_{0,0}(\omega)|$, all dendritic loci with an intermediate transfer impedance value can be mapped to a specific locus in the respective cylinder. This mapping guarantees (for the passive case) that the magnitude of the somatic voltage response, $V_0(\omega)$, to an input current, $I_X(\omega)$, injected at a dendritic location, X, will be identical between the full model and the reduced cylinder (see **Methods**). Although *Neuron_Reduce* is valid for any ω, conveniently, we found a close match between the full and the reduced models for ω = 0 (the steady-state case, where the transfer resistance is used rather than the transfer impedance, see **Figure 1D** and **Discussion**). Thus, all figures in this work are based on reduced models with ω = 0. Therefore, all the synapses and nonlinear ion channels are mapped to the respective locus in the reduced cylinder while preserving the transfer resistance to the soma (see **Figure 1** and **Methods**).

Neuro-reduce implemented on passive layer 5 pyramidal cell model with synapses

Figure 1 shows the case of implementing *Neuron_Reduce* on a detailed compartmental model of 3D reconstructed layer 5 pyramidal neuron from the rat somatosensory cortex

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Hay et al., 2011). This neuron consists of eight basal dendrites and one apical dendrite stemming from the soma, each of which is shown by a different color. This detailed model has active membrane ion channels at both soma and dendrites (see below). However, *Neuron_Reduce* first treats the modeled tree as passive by abolishing all membrane conductance while retaining only the leak conductance. Implementing Eqs. (1) - (11) for this cell (see **Methods**) produced a reduced, passive multi-cylindrical, model (**Figure 1B**, Step A). The detailed model consists of 642 compartments whereas the reduced model has only 48 compartments.

Figure 1C shows an example of four synapses located in different apical branches. These synapses all have the same transfer resistance to the soma. Therefore, they are all mapped to a single respective locus in the reduced cylinder, such that their transfer resistance is identical in the detailed and the reduced models. In the reduced model, these synapses are merged into one "NEURON" process (red synapse in **Figure 1B**) yet, they retain their individual activation time (see **Methods**). **Figure 1D** compares the transfer impedance between a specific point in the apical tree (marked by '**d**' in **Figure 1A** and **B**) and the soma. By construction, for the passive case, the transfer resistance (at $\omega = 0$) is identical in both models, validating *Neuron_Reduce* analytic method. Noteworthy is that, although constructed based on $\omega = 0$, the similarity between the full and reduced model is close also for higher input frequencies; it is somewhat larger in the reduced model for ω around 10 - 100 Hz (compare red to black lines). To test the performance of *Neuron_Reduce* for transient synaptic inputs (composed of mixed input frequencies), we activated the four synapses shown in **Figure 1C** sequentially in both the full and the reduced models (see **Methods** and **Table S2**). **Figure 1E** compared the composite somatic EPSP in the two models. The close similarity between the two EPSPs further validated that mapping the detailed model to the reduced model for $\omega = 0$ provides satisfactory results (see also **Figure S2**).

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

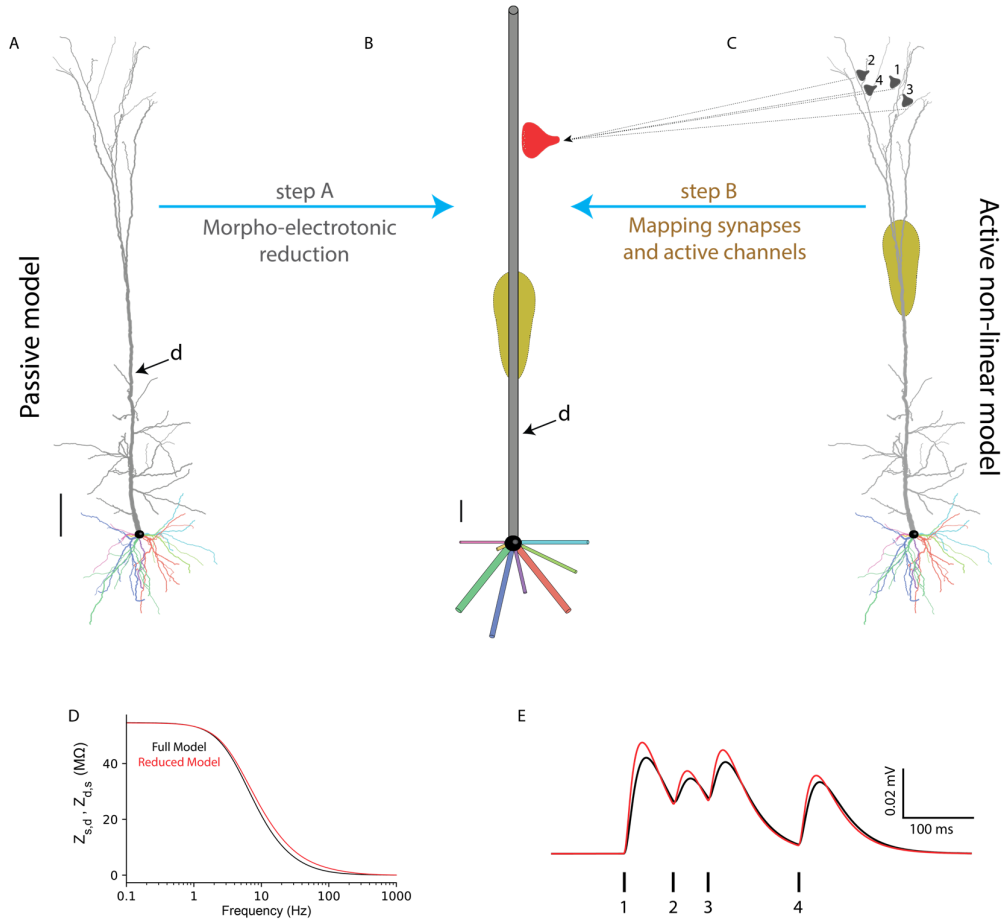

Figure 1. An analytic method for reducing neuron model complexity (*Neuron_Reduce*). **A**. Detailed passive model of 3D reconstructed L5 thick tufted pyramidal cell from rat neocortex; (the nine stem dendrites in the example shown are depicted by different colors). **B**. Each stem dendrite in the full model is reduced to a single cylinder that retains the specific passive cable properties (R_m, C_m, and R_i) of the original tree. The diameter and length of the respective cylinders are computed analytically using Eqs. (1) – (11), such that each cylinder preserves both the transfer resistance from the most distal dendritic tip to the soma as well as the input resistance at the soma end of the corresponding stem dendrite. This generates a unique cylindrical cable for each of the original stem dendrites. Scale bars in **A** and **B** are 100 µm. **C**. Synapses with the same transfer resistance to the soma (exemplar synapses are marked as 1-4 at top right) are all mapped to the respective locus in the reduced cylinder, such that their transfer resistance is identical in the two models. In the reduced model, these synapses are merged into one "NEURON" process (red synapse in **B**), yet they retain their individual activation time (see **Methods**). The same mapping also holds for active membrane conductances (schematic yellow region, denoting Ca^{2+} "hot spot" in the apical tree). **D**. Transfer impedance ($Z_{d,0} = Z_{0,d}$) between point **d** on the apical tree (shown in **A** and **B**) and the soma (X = 0) as a function of the input frequency in both the full (black trace) and the reduced model (red trace). **E**. Comparison of the composite somatic EPSP resulting from the sequential activation of the four synapses shown in **C** in the full model (black trace) and in the reduced model (red trace). The synapses were activated in temporal order 1, 2, 3, 4 (activation time is shown by the vertical lines below the composite EPSP; with respective peak conductances of 0.6, 0.3, 0.4 and 0.4 nS for AMPA-based synapses. See details in **Table S2**. In this example, both models have passive dendrites, see **Figure S1** for the active case.

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Accuracy and speed-up of *Neuron_Reduce* for a nonlinear neuron model

To measure the accuracy of *Neuron_Reduce* for a nonlinear neuron model we ran a comprehensive set of simulations using a well-established use-case, a L5 pyramidal cell model ((Hay et al., 2011), **Figure 2A** as in **Figure 1)**. This cell model also included several dendritic nonlinear channels including Ca^{2+} "hot spot" in the apical tuft (as in **Figure 1C**) and Na^{+} spiking mechanism in the cell body. We randomly distributed 8,000 excitatory and 2,000 inhibitory synapses on the modeled dendritic tree (synaptic parameters are listed in **Table S2**) and used *Neuron_Reduce* to generate a reduced model for this cell. Next, we simulated the full model by randomly activating the excitatory synapses at 5 Hz and the inhibitory synapses at 10 Hz (see **Methods**). The full model responded with an average firing rate of 11.8 Hz (black trace **Figure 2B**, only 2 sec out of 50 sec simulation time are shown). The average firing rate of the reduced model due to the same input was 11.1 Hz (red trace, **Figure 2B**; spike timings are shown by small dots on the top). The cross-correlation between the two spike trains peaked around zero (**Figure 2C**), and the inter-spike interval distributions were similar in the two models (**Figure 2D**).

The full range of responses to a random input for the two models was explored by varying the firing rate of the excitatory synapses and measuring the similarity between the spiking activity of the two models (**Figure 2E**). To assess the similarity between two spike trains we used the SPIKE-synchronization measure (Kreuz et al., 2015b, 2015a). We also tested other similarity metrics and found comparable results (not shown, see also Kreuz, 2011 and Satuvuori and Kreuz, 2018). The SPIKE-synchronization value of the two spike trains in **Figure 2B** is 0.785. The SPIKE-synchronization measure as a function of the synaptic firing rate is relatively low at small values, but for larger input rates it stabilizes around 0.75 (**Figure 2F**). A comparison of our method with two classical reduction methods, that of Rall d3/2 "equivalent cable" approach (Rall, 1962; Rall and Rinzel, 1973; Rinzel and Rall, 1974) and mapping (after cable filtering) all the synapses into a single somatic compartment (see **Methods**) are shown in **Figure S3**. In the entire range of inputs, *Neuron_Reduce* produced superior results. Furthermore, the model we have used in **Figure 2** consists of highly nonlinear dendrites (Hay et al., 2011). Decreasing the nonlinearity by using only AMPA-mediated synapses (without NMDA-component) further improved the similarity between the spike times in the reduced versus the full models (**Figure 2G**).

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

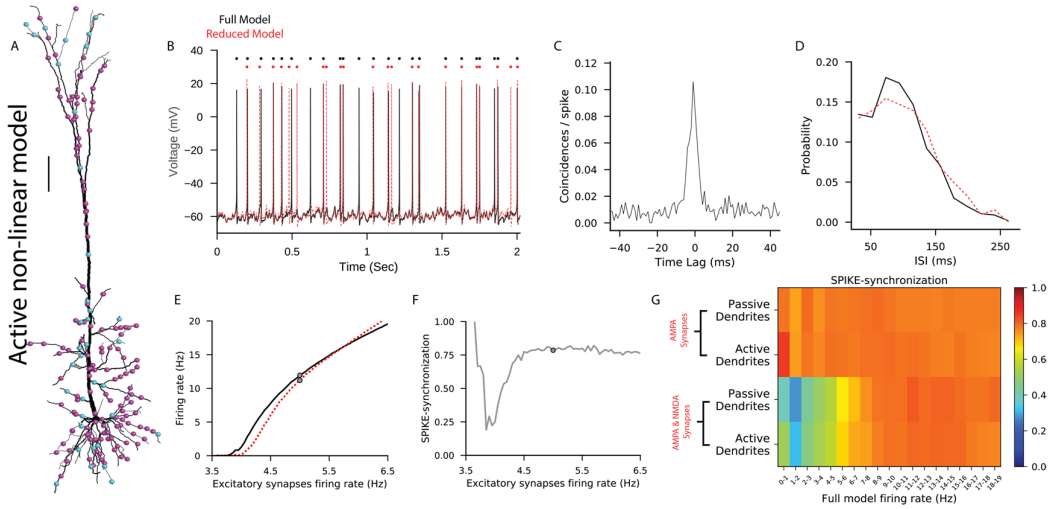

Figure 2. *Neuron_Reduce* faithfully replicated the I/O properties of a full nonlinear model of L5 pyramidal cell. A. Layer 5 pyramidal cell model (Hay et al., 2011) as in **Figure 1A**, with 8000 excitatory (magenta dots) and 2000 inhibitory synapses (cyan dots, see **Table S2** for synaptic parameters). Excitatory synapses were activated randomly at a rate of 5Hz and the inhibitory synapses at 10 Hz. This detailed model consists of dendritic Ca^{2+} "hot spot" (as in **Figure 1C**) and Na^{+} spiking mechanism at the cell body. Scale bar 100 µm. **B.** An example of the spiking activity at the soma of the full model (black trace) and of the reduced model (red trace); spike times are represented by the respective dots above the spikes. **C**. Cross-correlation between the spikes in the reduced versus the full models. **D**. Inter-Spike Interval (ISI) distributions for the two models. **E**. The output firing rate of the reduced (red) versus the full (black) models as a function of the firing rate of the excitatory synapses. **F.** SPIKE-synchronization measure (see **Methods**) between the two models as a function of the firing rate of the excitatory synapses. Grey dots in **E** and **F** represent the case shown in **B**. **G**. SPIKE-synchronization as a function of the biophysical and synaptic properties of the full model. The quality of the performance of the reduced versus the full model increases with input frequency and decreases with the addition of NMDA synapses.

In **Figure 3** we compared the run-time of the full versus the reduced models for the L5 cell shown in **Figure 2A**. Simulating the full model with 10,000 synapses for 50 seconds required 9,800 $\pm$ 3,500 seconds whereas in the reduced model it took only 130 $\pm$ 50 seconds, a ~80-fold improvement (see **Table S1**). The larger the number of synapses is in the full model, the longer the run-time is (**Figure 3A**). In contrast, the run-time in the reduced model is only slightly dependent on the number of synapses. This is expected when considering the synaptic merging step in our algorithm (see **Discussion**). The run-time in the reduced model does depend on the number of compartments per cylinder. However, although decreasing the number of compartments per cylinder decreases the faithfulness of the model, there was no improvement in the SPIKE-synchronization measure when ΔX per compartment was smaller than 0.1λ (**Figure 3B** and see also (Parnas and Segev, 1979)). Therefore, all the results presented in **Figures 1-5** are based on models with spatial discretization, ΔX, that is not larger than 0.1λ.

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

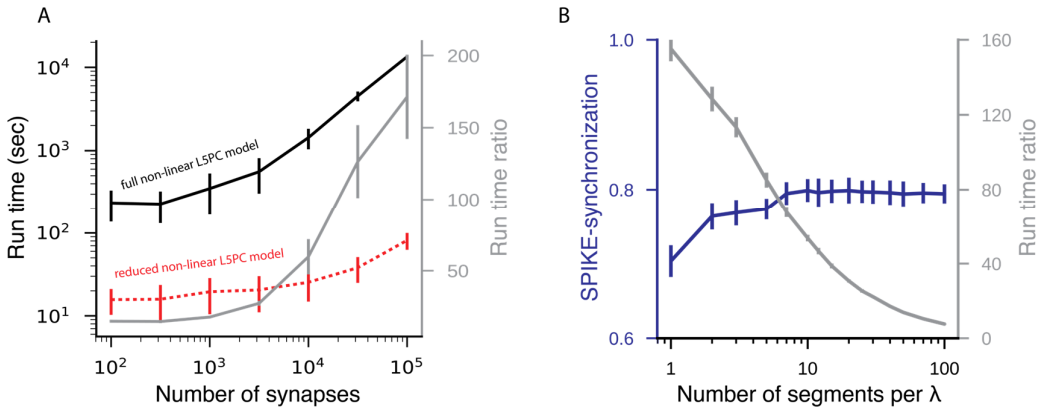

Figure 3. ***Neuron_Reduce*** **enhances the simulation speed by up to 200 folds. A**. Simulation run-time for the full (black) and the reduced models (red) of layer 5 pyramidal cell shown in **Figure 2A**, and their ratio (the speed-up, grey) as a function of the number of modeled synapses. **B**. Accuracy (blue) of the reduced models and their speed-up in simulation time (grey) as a function of the number of electrical segments per length constant (see also **Table S1**).

Neuron_Reduce replicated key dendritic nonlinearities and computations

To determine the capabilities of the reduced models to support nonlinear dendritic phenomena and dendritic computations, we repeated two classical experiments both in the full and the reduced model of the L5 pyramidal cell shown in **Figure 1**. The first simulated experiment starts with injecting a brief depolarizing step current to the soma of the full model in order to generate a somatic Na^+ action potential (AP, black trace in **Figure 4A**). This AP then propagated backward to the apical dendrite (red trace in **Figure 4A1**). Repeating the same current injection in the reduced model leads to a similar phenomenon (**Figure 4B1**). The full model also contains a "hot region" with voltage-dependent calcium conductances in its apical dendrite (in its nexus, see **Figure 1**). Combining the somatic current injection with synaptic-like transient depolarizing current injected to the apical nexus, evoked a prolonged Ca^{2+} spike in the distal apical dendrite (red trace at apical tree) which, in turn, generated a burst of somatic Na^+ spikes (the BAC firing, **Figure 4A2**, see (Hay et al., 2011; Larkum et al., 1999, 2009)). *Neuron_Reduce* maps this nonlinear dendritic "hot" Ca^{2+} region to its respective location in the reduced model (see **Figure 1** and **Methods**). **Figure 4B1-B2** shows that similar combination of somatic and dendritic currents produced the BAC firing phenomenon also in the reduced model. However, the reduced model is somewhat more excitable than the detailed model; this results in a burst of three spikes with higher frequency in the reduced model (compare **Figure 4A2** to **B2**).

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

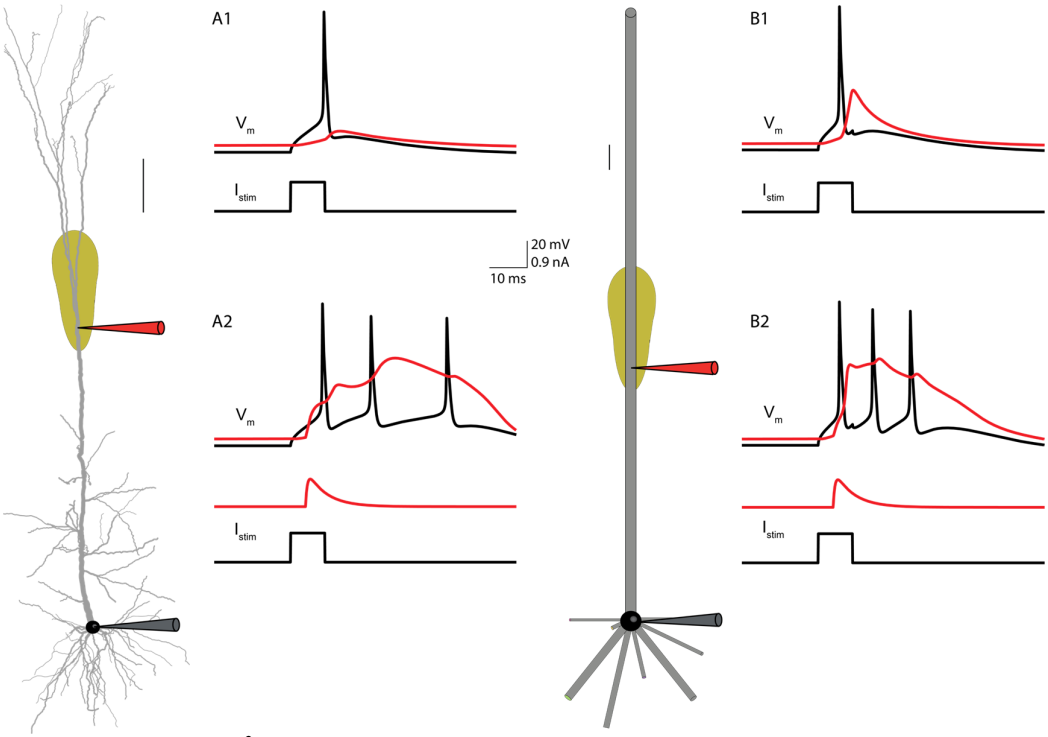

Figure 4. Dendritic Ca^{2+}-spike and BAC-firing are faithfully replicated in the reduced model. A1-A2 Left, the detailed L5 pyramidal cell model with nonlinear Ca^{2+} "hot spot" (same model as in **Figure 2** and **3**). **A1**. Injecting depolarizing step current to the soma (1.1 nA for duration of 11.9 ms) of the full model evoked a somatic action potential, AP (black trace) that propagated, semi-actively, backward into the apical tree (red trace). **A2**. Combining the somatic input with a transient synaptic-like current injection to the "hot region" in the apical dendrite evoked a prolonged local Ca^{2+} spike which, in turn, triggered a burst of two extra somatic Na^{+} spikes (the BAC firing phenomenon, Larkum et al., 1999). **B1-B2**. Back propagation and BAC firing in the reduced model with identical stimulations as in the full model. Synaptic current injections to the dendrite was of 0.95 nA peak value with 0.5 ms and 5 ms rise-time and decay-time respectively. Scale bar for the morphologies is 100 µm.

A second simulation attempted to replicate the original theoretical results of Rall (Rall, 1964) and (Anderson et al., 1999) and the experimental results of (Branco et al., 2010). In these studies, several excitatory synapses were distributed over a stretch of a basal dendrite and were activated sequentially in time, either in the soma-to-dendrites (IN) direction or *vice versa* (the OUT direction). Rall showed that the shape and size of the resultant composite somatic EPSP depends strongly on the spatio-temporal order of synaptic activation; it is always larger and more delayed for the centrifugal (soma-to-dendrites) than for the centripetal (dendrites-to-soma) sequence of synaptic activation (this difference could serve to compute the direction of motion; see (Anderson et al., 1999)). It was shown that the difference in the somatic voltage peak between these two spatio-temporal sequences of synaptic activation is enhanced when nonlinear NMDA-dependent synapses are involved (Anderson et al., 1999; Branco et al., 2010) and that it enables the discrimination between complex patterns of dendritic activation along the dendritic.

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To simulate these phenomena, twelve excitatory synapses were placed along one basal branch in the full model (red dots on the green basal tree, **Figure 5A1**). At first, the synapses had only an AMPA component. The synapses were activated in temporal order from the tip to the soma (IN, cyan traces) or from the soma to the tip (OUT, blue traces), see **Methods** for details. As predicted by Rall, the IN direction resulted with larger and delayed somatic EPSP (**Figure 5A2**). *Neuron_Reduce* merged these twelve synapses into five point processes along the respective cylinder (**Figure 5B1**). We repeated the same experiment in the reduced model and found that also in this model the resulted EPSP was larger and delayed for the IN direction, although the EPSPs waveform was not identical between the reduced and the full models (see **Figure S2** and **Discussion**). Next, an NMDA component was added to the 12 simulated synapses; this resulted in larger somatic EPSP amplitude in both directions (and both models) and a smaller difference in the peak timing between the different directions in both the full and the reduced model (compare **Figure 5A3** to **B3**).

To generalize the impact of the spatio-temporal order of synaptic activation a directionality index was suggested by (Branco et al., 2010). This measure estimates how different a given synaptic sequence is from the IN sequence by calculating the number of synaptic swaps needed in order to convert this given pattern into the IN pattern (using the bubble-sort algorithm, see **Methods**). We tested the EPSPs that resulted from different temporal combinations of synaptic activation that have a different directionality index, both without (**Figure 5C1**) and with NMDA component (**Figure 5D1**). The peak somatic EPSP in the reduced model (red dots) was larger than in the respective full model (black dots), both for the AMPA-only case (by 1.71 ± 0.43 mV), and for the AMPA + NMDA case (by 4.80 ± 0.74 mV); see **Figure S1**. Yet, when the somatic voltage was normalized by the peak voltage for the OUT direction, the behavior of the two models was similar (**Figures 5C2** and **5D2**). Now, the difference between the reduce and the full was only 0.11 ± 0.43 mV for the AMPA-only case and 0.35 ± 0.43 mV for the AMPA + NMDA. We conclude that, although the full and the reduced models differ, to a certain degree (mainly due to differences in the dendritic input resistances, see **Discussion**), the capability of the reduced model to discriminate between spatio-temporal patterns of synaptic activation is similar to that of the full model.

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

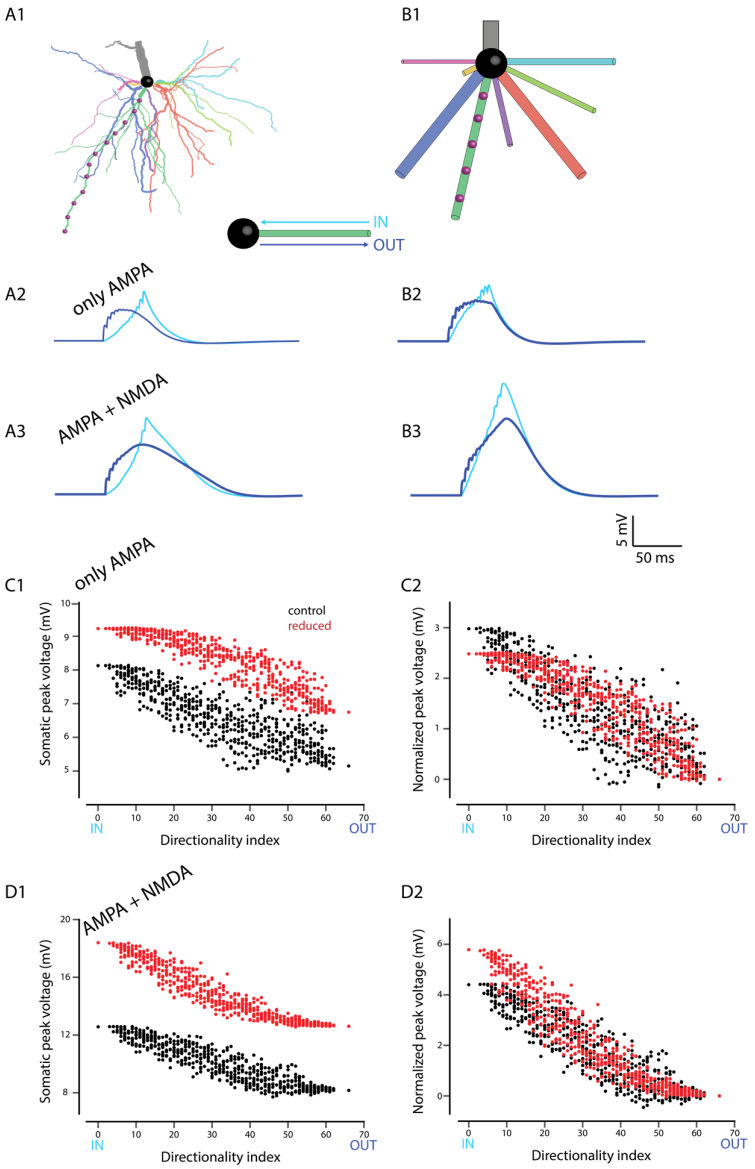

Figure 5. Sensitivity of a basal dendrite to spatio-temporal sequence of synaptic activation in the detailed versus the reduced models. A1. A model of L5PC (full model, **Figure 1**) with twelve excitatory synapses spatially distributed on one of its basal dendrites (red dots on green basal dendrite). **A2**. Somatic responses to sequential activations of its basal synapses in the IN (cyan) and the OUT (blue) directions. In this case the synaptic models consist of only AMPA component. **A3**. As is **A2** but the synaptic models consist of both AMPA and NMDA components. **B1**. Reduced model for the full model shown in **A1**. *Neuron_Reduce* mapped the twelve synapses in the full model into five synapses in the reduced model. **B2** and **B3**. As in **A2** and **A3** but for the reduced model. **C1**. Pattern separability (see **Methods**) of the full (black) and the reduced (red) models. **C2**. As in **C1** after substation of the peak voltage, obtained in the OUT direction, from each of the voltage responses. **D1** and **D2**. As in **C1** and **C2** but when the synaptic models consist of both AMPA and NMDA conductances. Note the similarity between the full and the reduced model in terms of pattern separability.

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Neuron-Reduce was implemented successfully on a variety of neuron models

Next, we tested the utility of *Neuron_Reduce* for eight different neuron models (**Figure 6**). These neuron models were extracted from three different databases; each model is for a different cell type. Four models were taken from the Blue Brain database (Markram et al., 2015) - L6 tufted pyramidal cell, L4 double bouquet cell, L4 spiny stellate cell and L5 Martinotti cell from the rat somatosensory cortex. Two additional models were taken from the Allen Institute cell type database (Gouwens et al., 2018) – an L4 spiny cell and L1 aspiny cell from the mouse visual cortex. Two additional neuron models are coming from our lab, a model of a rat L2/3 large basket cell (Amsalem et al., 2016) and a model of a human L2/3 PC (Eyal et al., 2016). All these models were constrained by experimental recordings and 3D reconstructions. See **Table S2** for details about the various neuron models and input parameters.

Although the models are for different neuronal types, have different morphologies, and different electrical properties. *Neuron_Reduced* managed to successfully generate a respective reduced model with highly faithful response properties for each of them (**Figure 6**). Three examples with the morphologies of the full and the reduced models are shown in **Figure 6A-C**. For a given input we measured the spiking activity of the full and the reduced models (**Figure 6D-F)** and calculated the corresponding SPIKE-synchronization values. For the L6 tufted PC model **(Figures 6A,D)**, the L2/3 large basket cell model **(Figures 6B,E)**, and the L6 double bouquet model **(Figures 6C,F)** the SPIKE-synchronization values were 0.74, 0.85, and 0.91 respectively, for a trace of 50 seconds (only 2 seconds are shown in **Figure 6**). The spike train similarities (in SPIKE-synchronization) for other inputs and for the other five neuron models (and their corresponding reduced models) are shown in **Figure 6G**.

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

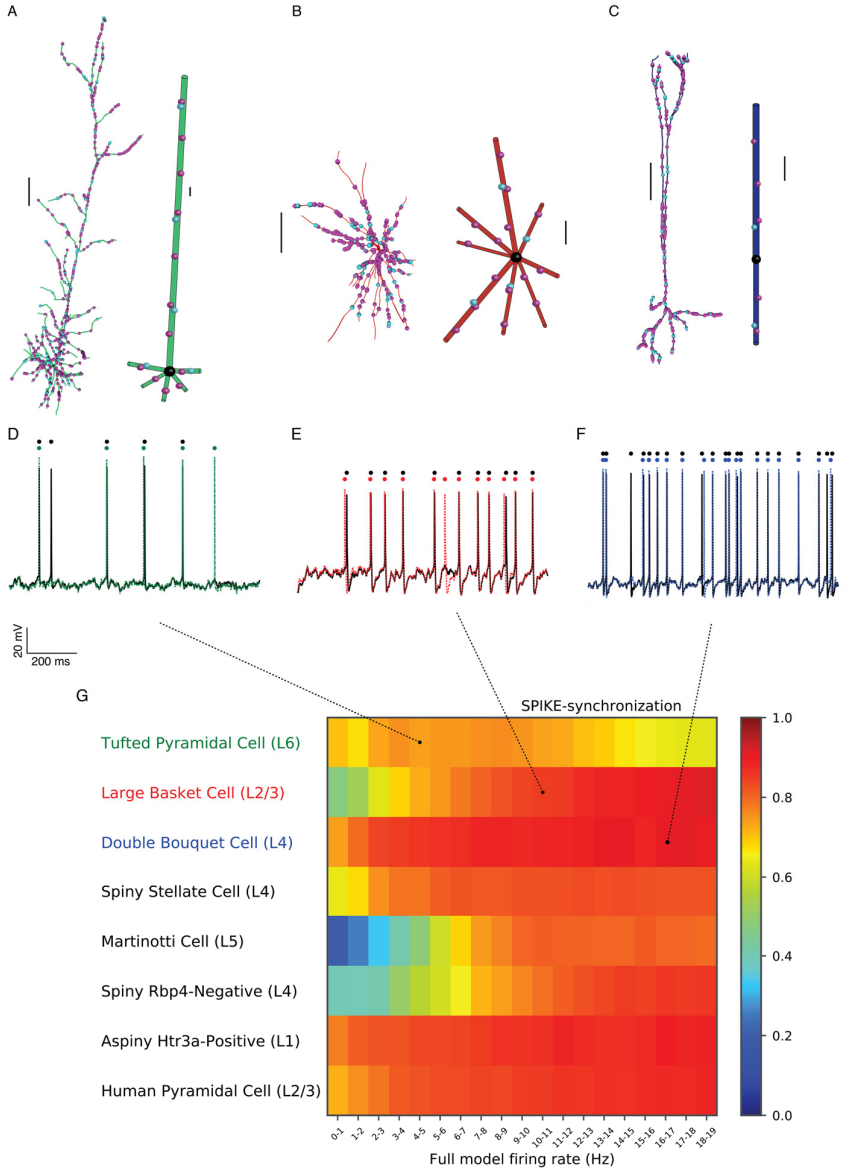

Figure 6. ***Neuron_Reduce*** **working successfully on a variety of neuron models. A-C**. Detailed models of three somatosensory neurons (left, L6 tufted pyramidal cell in green, middle, L2/3 large basket cell in red, and right, L4 double bouquet cell in blue) and their respective reduced models. Scale bars 100 µm. **D-F**. Voltage responses to an excitatory synaptic input activated at 1.2, 1.2, 2.7 Hz respectively of both the full (black) and the reduced models (corresponding colors). The inhibitory input activation rate was 10 Hz for all models. SPIKE-synchronization index for the full versus the reduced for eight neuron models. See **Table S2** for cell models and input parameters.

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

In this work, we presented *Neuron_Reduce*, a new tool for simplifying complex neuron models and for enhancing their simulation run-time. This tool maps the full tree analytically into a reduced multi-cylindrical tree, based on Rall's cable theory and linear circuit theory (**Figure 1**). The underpinning of the reduction algorithm is that it preserves (analytically) the magnitude of the transfer impedance $|Z_{0,j}|$ from each dendritic location, j, to soma (the dendro-somatic direction, Eqs. 1-11). Because in linear systems $|Z_{0,j}| = |Z_{j,0}|$, it also preserves, for passive dendritic trees, the transfer resistance in the somato-dendritic direction (e.g., current injection in or near the soma will result in the same voltage response at the respective sites in the full and reduced models; see also Koch et al., 1982). *Neuron_Reduce* is straightforward to use; it is fast, accurate and general so that it could be implemented on any neuron's morphology with a realistic number (tens of thousands) of synapses per neuron. One key advantage of *Neuron_Reduce* is that retains the identity of individual dendrites and synapses and it maps dendritic nonlinearities to their respective loci in the reduced model, thus preserving local excitable dendritic phenomena, thus maintaining certain dendritic computations. It also preserves the passive cable properties (R_m, R_a, *and* C_m) of the full model, thus preserving synaptic integration and other temporal aspects. *Neuron_Reduce* enhances the computational speed by a factor of up to 200, depending on the simulated morphology and the number of simulated synapses (**Figures 3** and **Table S1**). These mixtures of capabilities, together with its user-friendly documentation and its public availability makes *Neuron_Reduce* a favorable method for the community of neuronal modelers and computational neuroscientists, and for the growing community interested in "biophysical deep learning".

Several other reduction methods for single neurons have been proposed along the years (Bahl et al., 2012; Bush and Sejnowski, 1993; Destexhe, 2001; Hendrickson et al., 2011; Marasco et al., 2012; Rössert et al., 2016; Stratford et al., 1989). Most of them are not based on analytic underpinning and, thus, require hand-tuning of the respective biophysical and morphological parameters. Most of these methods were not examined using realistic numbers of dendritic synapses and are incapable of incorporating systematically dendritic nonlinearities and, in most cases, their accuracy was not assessed for a variety of neuron types (but see Marasco et al., 2013). Many of these methods are not well-documented and, therefore, it is hard to directly compare them

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with our method. Nevertheless, we did compare the performance of *Neuron_Reduce* to that of two other reduction methods and showed the advantage of *Neuron_Reduce* over these methods (**Figure S3**).

The reduced model for nonlinear layer 5 pyramidal cells reproduced local dendritic spikes (**Figure 4**) and certain dendritic computations (**Figure 5**). However, it is important to emphasize that, although the transfer impedance from a given dendritic loci to the soma is preserved, the input impedance at that loci is not preserved in the reduced model. Consequently, the conditions for evoking local dendritic events, and the fine details of these events, are not necessarily identical in the full and the reduced models (e.g., compare **Figure 4A1** and **A2** to **Figure 4B1** and **B2**). Similarly, because the local voltage response to a current injection in the dendrite depends on the dendritic impedance, the local synaptic responses is somewhat different in the two cases, especially when voltage-gated ion channels (such as NMDA-dependent synaptic channels) are involved. Indeed, when large NMDA signals are involved the resultant somatic EPSP are expected to be different between the full and the reduced model as is the case in **Figures 4** and **5**. Albeit these local differences, the models generated by *Neuron_Reduce* were capable of generating local dendritic Ca^{2+} spike in the cylinder representing the apical dendrite as well as performing input classification task (enhanced by NMDA-conductance) as in the full tree (**Figures 4** and **5**). We note that, when embedded in large circuits, individual neurons are likely to receive semi-random dendritic input rather than a clustered input on specific dendrites. For such inputs, the reduced models generated by *Neuron_Reduce* captures most of the statistics of the membrane voltage dynamics as in the full model (**Figures 2** and **6**).

For a large number of synapses and complex morphologies, the run-time of *Neuron_Reduce* models could be accelerated by up to 200 folds (**Figure 3** and **Table S1**). This is achieved by two steps that are linked together. First, the algorithm reduces the number of compartments per neuron model; e.g., for the reconstructed tree in **Figure 1**, it reduced the number of compartments from 642 to 48 . Then synapses (and ion channel) that are mapped to the same electrical compartment in the reduced tree (because they have similar transfer resistance to the soma) are merged to one point process in NEURON. Each one of these steps on its own has a relatively small effect on the run-time. However, when combining the two, a large (supra-linear) improvement in the computational speed is achieved (**Table S1**). This is because, in each time step,

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

NEURON computes both the voltage in each electrical compartment as well as the currents and states of each point process and membrane mechanism (synapses and conductances). Reducing the number of compartments in a model both decreases the number of equations to be solved and the number of synapses to be simulated (due to the reduced number of compartments larger number of synapses are merged together). Importantly, merging synapses preserves the activation time of each synapse. We note, however, that at its present state, *Neuron_Reduce* cannot merge synapses with different kinetics.

The next straightforward step is to use *Neuron_Reduce* for simplifying all the neurons composing a large neural network model (as in the Blue Brain Project (Markram et al., 2015)), and in the *in silico* model of (Egger et al., 2014). By keeping the connectivity, and reducing the complexity of the neuronal models, the reduced models will allow running longer simulations and/or larger neuronal networks, while faithfully preserving the I/O of each neuron. Such long simulations are critical for reproducing long-term processes such as circuit evolution and structural and functional plasticity.

Methods

Neuron_Reduce algorithm and its implementation in NEURON

Neuron_Reduce transforms each stem dendrite in the full model to a unique single cylinder with both ends sealed. This cylinder preserves the specific passive cable properties (R_m, C_m, and R_i) of the original tree and both the transfer resistance obtained from the electrotonically most distal dendritic tip to the soma and the input resistance at the soma end of the corresponding stem dendrite (when disconnected from the soma). For a sinusoidal angular frequency $\omega > 0$, the transfer impedance $Z_{i,j}(\omega)$ is the ratio between the Fourier transform of the voltage in the measured point (i) and the sinusoidal current injected in the injection point (j) (note that, in a passive system, $Z_{i,j}(\omega) = Z_{j,i}(\omega)$). This ratio is a complex number; its magnitude ($|Z_{i,j}(\omega)|$) is the ratio (in Ω) between the peak voltage response and the amplitude of the injected current where ϕ is the phase difference between the sinusoidal input current and the sinusoidal voltage response. In a short cylindrical cable with sealed ends and electrotonic length L, the transfer impedance, $Z_{0,X}(\omega)$, between the somatic end of the cylinder (X = 0) and any location X is (Rall and Segev, 1985; Koch, 1999):

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$Z_{0,X}(\omega) = \frac{R_\infty}{q}\frac{cosh(q(L-X))}{sinh(qL)} \tag{1}$$

where

$$R_\infty = \frac{2}{\pi}\frac{\sqrt{R_m R_i}}{d^{3/2}} \tag{2}$$

and

$$q = \sqrt{1 + i\omega\tau} \tag{3}$$

and τ is the membrane time constant $R_m C_m$.

From Eq. (1), the input impedance at X = 0 is,

$$Z_{0,0}(\omega) = \frac{R_\infty}{q}\coth(qL) \tag{4}$$

We seek for a cable of length L in which both $|Z_{0,L}(\omega)|$ and $|Z_{0,0}(\omega)|$ are identical to those measured in the respective stem dendrite in the full model (**Figure 1**). Towards this end we first search for an L value in which the ratio $|Z_{0,L}(\omega)|/|Z_{0,0}(\omega)|$ is preserved. Dividing Eq. (1) by Eq. (4):

$$\frac{Z_{0,X}(\omega)}{Z_{0,0}(\omega)} = \frac{\cosh(q(L-X))}{\cosh(qL)} \tag{5}$$

which could be expressed as,

$$\frac{Z_{0,X}(\omega)}{Z_{0,0}(\omega)} = \frac{\cosh(a(L-X)+ib(L-X))}{\cosh(aL+ibL)} = \mathrm{M}exp(i\phi) \tag{6}$$

where a and b are the real and the imaginary parts of q, respectively, and M and ϕ are the modulus and phase angle of this complex ratio.

From (Rall and Segev, 1985) it follows that,

$$\mathrm{M} = \frac{|Z_{0,X}(\omega)|}{|Z_{0,0}(\omega)|} = \left[\frac{\cosh(2a(L-X))+\cos(2b(L-X))}{\cosh(2aL)+\cos(2bL)}\right]^{0.5} \tag{7}$$

and

$$\phi = \arctan\left[\tanh\left(a(L-X)\right)\tan\left(b(L-X)\right)\right] - \arctan[\tanh(aL)\tan(bL)] \tag{8}$$

Importantly, for a fixed M (and a given ω) there is a unique value for L that satisfies Eq. (7) (see **Figure 4** in (Rall and Segev, 1985) and note the-one-to-one mapping between M and L for a given ω value). However, there are infinite number of cylindrical cables (with different diameter and length) with identical L values that preserve a given M value in Eq. (7).

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We, next seek for a unique cable, with diameter d, that also preserves $|Z_{0,0}(\omega)|$ (and therefore also $|Z_{0,L}(\omega)|$ see Eq. (7)).

From Eqs. (2) and (4) we get,

$$Z_{0,0}(\omega) = \frac{2}{\pi q}\frac{\sqrt{R_m R_i}}{d^{3/2}}\coth(qL) \tag{9}$$

And thus,

$$\left|Z_{0,0}(\omega)\right| = \left|\frac{2}{\pi q}\frac{\sqrt{R_m R_i}}{d^{3/2}}\coth(qL)\right| \tag{10}$$

From which we compute the diameter, *d*, for that cylinder,

$$|d| = \left|\left(\frac{2}{\pi}\frac{\sqrt{R_m R_i}}{qZ_{0,0}(\omega)}\coth(qL)\right)^{2/3}\right| \tag{11}$$

Eqs. (1) – (11) provides the unique cylindrical cable (with a given d, L, and specific membrane and axial properties) that preserves the values of $\left|Z_{0,L}(\omega)\right|$ and $\left|Z_{0,0}(\omega)\right|$ that were measured in the respective stem dendrite.

Note that this unique cable does not necessarily preserve the phase ratio (ϕ in Eq. (8)) as in the original tree.

Practically, in order to transform each original stem dendrite (with fixed R_m, R_i, *and* C_m) to a corresponding unique cylindrical cable, we proceeded as follows: First, we searched, for each modeled stem dendrite (when isolated from the soma), the distal location *X* with the minimal transfer impedance $\left|Z_{0,X}\right|$ from *X* to the soma. Therefore, this location provides the smallest M value for this particular stem dendrite. This distal dendritic locus is mapped to the distal end, X = L, of the corresponding cylinder. We then used Eqs. (1) – (11) to calculate the unique cylinder for that stem dendrite. Next, for any intermediate location, *x*, in the stem dendrite, we find the respective location in the cylinder that preserves $\left|Z_{0,x}(\omega)\right|$. Note that $\left|Z_{0,0}(\omega)\right|$, $\left|Z_{0,L}(\omega)\right|$, could be computed analytically for the original stem dendrite using NEURON impedance tool (Carnevale et al., 1995) and Eq. (7), and Eq. (11) above.

Merging the synapses was achieved using the Point Process object in NEURON. All synapses with the same $\left|Z_{0,x}(\omega)\right|$ value were mapped to the respective location in the

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

reduced model and were merged to a one point process (**Figure 1**, step B). These synapses retain their original activation time and biophysical properties by connecting this point process to the original NetCon objects for the respective synapses. As shown in **Table S1** this step dramatically reduces the running time of the model. All the results in this study use $\omega = 0$ in Eqs. (1-11), as running the same simulations with other values of ω did not improve the match between the full and the reduced models (not shown). However, ω is a parameter in the algorithm code and can be modified by the user.

Neuron models used in the present study

To estimate the accuracy of the reduction method we ran simulations for 50 seconds in both the reduced and the full models. Models of 12 neurons were used in this study, and their details are available in **Table S2**. For each of the models, we distributed 1,250 - 10,000 synapses on their dendritic trees. Eighty percent of the synapses were excitatory, and the rest were inhibitory. The synaptic conductances were modeled using two-state kinetic synaptic models, based on the synaptic models in (Markram et al., 2015). For simplicity, we did not include synaptic facilitation or depression. All models had one type of inhibitory synapses ($GABA_A$), and one or two types of excitatory synapses (AMPA and NMDA). The synaptic rise and decay time constants were taken from various works cited in **Table S2**. When no data was available, we used the default parameters of the BBP project synaptic models. Inhibitory synapses were activated at 10 Hz, whereas the activation rate of the excitatory synapses was varied to generate different output firing rates, in the range of 1 to 20 Hz (**Figures 2** and **6**); the values used for each model is available in **Table S2**.

Estimating the accuracy of the reduced models

Cross-correlation was measured between the spike trains of the full and the reduced models. The window size was 500 ms, and the bin size was 1 ms. The resulted cross-correlation was normalized by the number of spikes in the full model (**Figure 2C**). Interspike intervals (**ISI**) were binned in windows of 21 ms to create the ISI distribution in **Figure 2D**.

SPIKE-synchronization measure (**Figures, 2F**, **G**, **Figure 3B, Figure 6** and **Figure S3**) is a parameter- and scale-free method that quantifies the degree of synchrony between two spike trains (Kreuz et al., 2015b). SPIKE-synchronization uses the relative number

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of quasi-simultaneous appearances of spikes in the spike trains. In this study, we used the Python implementation of this method (Mulansky and Kreuz, 2016).

Comparison with other reduction algorithms

We compared *Neuron_Reduce* to two classical reduction algorithms (**Figure S3**):

*1. Equivalent cable with d3/2 rule***,** Rall and Rinzel (Rall and Rinzel, 1973; Rinzel and Rall, 1974) showed that, for idealized passive dendritic trees it is possible to collapse the entire dendritic tree to a single cylinder (Rall's "equivalent cylinder"), that is analytically identical to the full tree. However, neurons do not have idealized dendritic trees, mostly because dendritic terminations typically occur at different electrical distances from the soma for different subtrees. Still it is possible to collapse any dendritic tree using similar mapping ("d3/2 rule") as in the idealized tree; this will provide an "equivalent cable" (rather than "equivalent cylinder") with a varying diameter for the whole dendritic tree (see details in Rall et al., 1992). The electrotonic distance between synapses and non-linear dendritic mechanisms were computed for the original model and then mapped to the corresponding segment in the "equivalent cable" having the same electrotonic distance to the soma as in the original tree.

*2. Mapping all synapses to the soma***.** Another common reduction scheme is to map all dendritic synapses to a single "somatic" compartment (Rössert et al., 2016). Here we used a modified version of this method. We used *Neuron_Reduce* to generate a multi-cylindrical model of this cell as in **Figure 1B**. Then all synapses in the original tree were mapped to the soma of that model. To account for the dendritic filtering, we multiplied the original synaptic conductance, g_{syn}, by the steady-state voltage attenuation factor from the original dendritic location, j, of that synapse to the soma. Namely,

$$g_{syn}^{*} = g_{syn} * \frac{|Z_{0,j}|}{|Z_{0,0}|} = g_{syn} * \frac{V_{0,j}}{V_{0,0}} \qquad (13)$$

where g_{syn}^{*} is the new synaptic weight for synapse j when placed at the soma of the reduced model.

Discriminating between spatio-temporal patterns of synaptic activation.

In **Figure 5** twelve synapses, 25 µm apart from each other, were distributed on a stretch of one basal dendrite. The peak AMPA conductance per synapse was 5 nS. In cases where the synapses also had an NMDA component, its conductance was 3.55 nS. The

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

synapses were activated in specific temporal order with a time delay of 3.66 ms between them, such that all the synapses were activated within 40.3 ms. This resulted in input velocity of 7 μm/s for the sequential patterns IN and OUT in **Figure 5**. In addition, the temporal order of synaptic activation was randomized and scored according to the directionality index (Branco et al., 2010), which sums how many swaps are being used by the bubble sort algorithm in order to sort specific pattern into the IN pattern. An IN pattern gets the value of 0 (no swaps) whereas the OUT pattern gets the value of 67 (67 swaps in bubble sort as in Branco et al., 2010)

Simulations were performed using the NEURON 7.4 (Carnevale and Hines 2006) running on a local cluster of 40 Intel(R) Xeon(R) CPU E5-2670 with 16 cores per node (640 cores in total)

Neuron_Reduce algorithm is publicly available in GitHub (http://github.com/orena1/neuron_reduce).

Acknowledgment
This research received funding from the European Union's Horizon 2020 Framework Program for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2) and from the Gatsby charitable foundation.

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Aamir, S. A., Muller, P., Hartel, A., Schemmel, J., and Meier, K. (2016). A highly tunable 65-nm CMOS LIF neuron for a large scale neuromorphic system. in *ESSCIRC Conference 2016: 42nd European Solid-State Circuits Conference* (IEEE), 71–74. doi:10.1109/ESSCIRC.2016.7598245.

Amsalem, O., Van Geit, W., Muller, E., Markram, H., and Segev, I. (2016). From Neuron Biophysics to Orientation Selectivity in Electrically Coupled Networks of Neocortical L2/3 Large Basket Cells. *Cereb. Cortex* 26, 3655–3668. doi:10.1093/cercor/bhw166.

Anderson, J. C., Binzegger, T., Kahana, O., Martin, K. A. C., and Segev, I. (1999). Dendritic asymmetry cannot account for directional responses of neurons in visual cortex. *Nat. Neurosci.* 2, 820.

Bahl, A., Stemmler, M. B., Herz, A. V. M., and Roth, A. (2012). Automated optimization of a reduced layer 5 pyramidal cell model based on experimental data. *J. Neurosci. Methods* 210, 22–34. doi:10.1016/j.jneumeth.2012.04.006.

Bower, J. M. (1998). "Constructing new models," in *The Book of GENESIS* (Springer), 195–201.

Branco, T., Clark, B. A., and Häusser, M. (2010). Dendritic discrimination of temporal input sequences in cortical neurons. *Science (80-.).* 329, 1671–1675.

Brown, S. A., Moraru, I. I., Schaff, J. C., and Loew, L. M. (2011). Virtual NEURON: A strategy for merged biochemical and electrophysiological modeling. *J. Comput. Neurosci.* 31, 385–400. doi:10.1007/s10827-011-0317-0.

Bush, P. C., and Sejnowski, T. J. (1993). Reduced compartmental models of neocortical pyramidal cells. *J. Neurosci. Methods* 46, 159–166. doi:10.1016/0165-0270(93)90151-G.

Cantarelli, M., Marin, B., Quintana, A., Earnshaw, M., Court, R., Gleeson, P., et al. (2018). Geppetto: a reusable modular open platform for exploring neuroscience data and models. *Philos. Trans. R. Soc. B Biol. Sci.* 373, 20170380. doi:10.1098/rstb.2017.0380.

Carnevale, N. T., and Hines, M. L. (2006). *The NEURON Book*. Available at: https://books.google.co.il/books/about/The_NEURON_Book.html?id=YzcOyjKBPHgC&pgis=1 [Accessed July 30, 2015].

Carnevale, N. T., Tsai, K. Y., Claiborne, B. J., and Brown, T. H. (1995). "The Electrotonic Transformation: a Tool for Relating Neuronal Form to Function," in *Advances in Neural Information Processing Systems 7*, eds. G. Tesauro, D. S. Touretzky, and T. K. Leen (MIT Press), 69–76. Available at: http://papers.nips.cc/paper/945-the-electrotonic-transformation-a-tool-for-relating-neuronal-form-to-function.pdf.

Destexhe, A. (2001). Simplified models of neocortical pyramidal cells preserving somatodendritic voltage attenuation. *Neurocomputing* 38–40, 167–173. doi:10.1016/S0925-2312(01)00428-3.

Egger, R., Dercksen, V. J., Udvary, D., Hege, H.-C., and Oberlaender, M. (2014). Generation of dense statistical connectomes from sparse morphological data. *Front. Neuroanat.* 8, 129. doi:10.3389/fnana.2014.00129.

Eyal, G., Verhoog, M. B., Testa-Silva, G., Deitcher, Y., Lodder, J. C., Benavides-Piccione, R., et al. (2016). Unique membrane properties and enhanced signal processing in human neocortical neurons. *Elife* 5, e16553.

Gleeson, P., Steuber, V., and Silver, R. A. (2007). neuroConstruct: a tool for modeling networks of neurons in 3D space. *Neuron* 54, 219–235.

Gouwens, N. W., Berg, J., Feng, D., Sorensen, S. A., Zeng, H., Hawrylycz, M. J., et al. (2018). Systematic generation of biophysically detailed models for diverse

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cortical neuron types. *Nat. Commun.* 9, 710.

Hawrylycz, M., Anastassiou, C., Arkhipov, A., Berg, J., Buice, M., Cain, N., et al. (2016). Inferring cortical function in the mouse visual system through large-scale systems neuroscience. *Proc. Natl. Acad. Sci.* 113, 7337–7344.

Hay, E., Hill, S., Schürmann, F., Markram, H., and Segev, I. (2011). Models of neocortical layer 5b pyramidal cells capturing a wide range of dendritic and perisomatic active properties. *PLoS Comput. Biol.* 7, e1002107. doi:10.1371/journal.pcbi.1002107.

Hendrickson, E. B., Edgerton, J. R., and Jaeger, D. (2011). The capabilities and limitations of conductance-based compartmental neuron models with reduced branched or unbranched morphologies and active dendrites. *J. Comput. Neurosci.* 30, 301–21. doi:10.1007/s10827-010-0258-z.

Koch, C. (1999). *Biophysics of computation: information processing in single neurons*. Oxford university press.

Koch, C., Poggio, T., and Torres, V. (1982). Retinal Ganglion Cells: A Functional Interpretation of Dendritic Morphology. *Philos. Trans. R. Soc. B Biol. Sci.* 298, 227–263. doi:10.1098/rstb.1982.0084.

Kozloski, J., and Wagner, J. (2011). An Ultrascalable Solution to Large-scale Neural Tissue Simulation. *Front. Neuroinform.* 5, 15. doi:10.3389/fninf.2011.00015.

Kreuz, T. (2011). Measures of spike train synchrony. *Scholarpedia* 6, 11934. doi:10.4249/scholarpedia.11934.

Kreuz, T., Bozanic, N., and Mulansky, M. (2015a). SPIKE-Synchronization: a parameter-free and time-resolved coincidence detector with an intuitive multivariate extension. *BMC Neurosci.* 16, P170. doi:10.1186/1471-2202-16-S1-P170.

Kreuz, T., Mulansky, M., and Bozanic, N. (2015b). SPIKY: a graphical user interface for monitoring spike train synchrony. *J. Neurophysiol.* 113, 3432–3445. doi:10.1152/jn.00848.2014.

Larkum, M. E., Nevian, T., Sandler, M., Polsky, A., and Schiller, J. (2009). Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. *Science* 325, 756–60. doi:10.1126/science.1171958.

Larkum, M. E., Zhu, J. J., and Sakmann, B. (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature* 398, 338–341. doi:10.1038/18686.

Magee, J. C., and Cook, E. P. (2000). Somatic EPSP amplitude is independent of synapse location in hippocampal pyramidal neurons. *Nat. Neurosci.* 3, 895.

Marasco, A., Limongiello, A., and Migliore, M. (2012). Fast and accurate low-dimensional reduction of biophysically detailed neuron models. *Sci. Rep.* 2, 928. doi:10.1038/srep00928.

Marasco, A., Limongiello, A., Migliore, M., Vogels, T. P., Rajan, K., Abbott, L. F., et al. (2013). Using Strahler's analysis to reduce up to 200-fold the run time of realistic neuron models. *Sci. Rep.* 3, 2934. doi:10.1038/srep02934.

Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., et al. (2015). Reconstruction and Simulation of Neocortical Microcircuitry. *Cell* 163, 456–492.

Migliore, M., Hoffman, D. A., Magee, J. C., and Johnston, D. (1999). Role of an A-type K+ conductance in the back-propagation of action potentials in the dendrites of hippocampal pyramidal neurons. *J. Comput. Neurosci.* 7, 5–15.

Mulansky, M., and Kreuz, T. (2016). PySpike—A Python library for analyzing spike train synchrony. *SoftwareX* 5, 183–189.

Parnas, I., and Segev, I. (1979). A mathematical model for conduction of action

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

potentials along bifurcating axons. *J. Physiol.* 295, 323–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/521942 [Accessed December 6, 2018].

Poirazi, P., Brannon, T., and Mel, B. W. (2003). Arithmetic of subthreshold synaptic summation in a model CA1 pyramidal cell. *Neuron* 37, 977–987.

Rall, W. (1962). Electrophysiology of a Dendritic Neuron Model. *Biophys. J.* 2, 145–167. doi:10.1016/S0006-3495(62)86953-7.

Rall, W. (1964). Theoretical significance of dendritic trees for neuronal input-output relations. *Neural Theory Model.*, 73–97.

Rall, W. (1967). Distinguishing theoretical synaptic potentials computed for different soma-dendritic distributions of synaptic input. *J. Neurophysiol.* 30, 1138–68.

Rall, W., Burke, R. E., Holmes, W. R., Jack, J. J. B., Redman, S. J., and Segev, I. (1992). Matching dendritic neuron models to experimental data. *Physiol. Rev.* 72.

Rall, W., and Rinzel, J. (1973). Branch input resistance and steady attenuation for input to one branch of a dendritic neuron model. *Biophys. J.* 13, 648–87. doi:10.1016/S0006-3495(73)86014-X.

Rall, W., and Segev, I. (1985). "Space-clamp problems when voltage clamping branched neurons with intracellular microelectrodes," in *Voltage and patch clamping with microelectrodes* (Springer), 191–215.

Rapp, M., Yarom, Y., and Segev, I. (1996). Modeling back propagating action potential in weakly excitable dendrites of neocortical pyramidal cells. *Proc. Natl. Acad. Sci.* 93, 11985–11990.

Rinzel, J., and Rall, W. (1974). Transient response in a dendritic neuron model for current injected at one branch. *Biophys. J.* 14, 759–790.

Rössert, C., Pozzorini, C., Chindemi, G., Davison, A. P., Eroe, C., King, J., et al. (2016). Automated point-neuron simplification of data-driven microcircuit models. *arXiv Prepr. arXiv1604.00087*.

Satuvuori, E., and Kreuz, T. (2018). Which spike train distance is most suitable for distinguishing rate and temporal coding? *J. Neurosci. Methods* 299, 22–33. doi:10.1016/J.JNEUMETH.2018.02.009.

Schemmel, J., Fieres, J., and Meier, K. (2008). Wafer-scale integration of analog neural networks. in *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on* (IEEE), 431–438.

Segev, I. (1992). Single neurone models: oversimple, complex and reduced. *Trends Neurosci.* 15, 414–421. doi:10.1016/0166-2236(92)90003-Q.

Segev, I., and London, M. (1999). A theoretical view of passive and active dendrites. *Dendrites, pages xxi* 376.

Stratford, K., Mason, A., Larkman, A., Major, G., and Jack, J. (1989). "The Computing Neuron," in, eds. R. Durbin, C. Miall, and G. Mitchison (Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.), 296–321. Available at: http://dl.acm.org/citation.cfm?id=103938.103954.

Stuart, G., and Spruston, N. (1998). Determinants of voltage attenuation in neocortical pyramidal neuron dendrites. *J. Neurosci.* 18, 3501–3510.

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Information

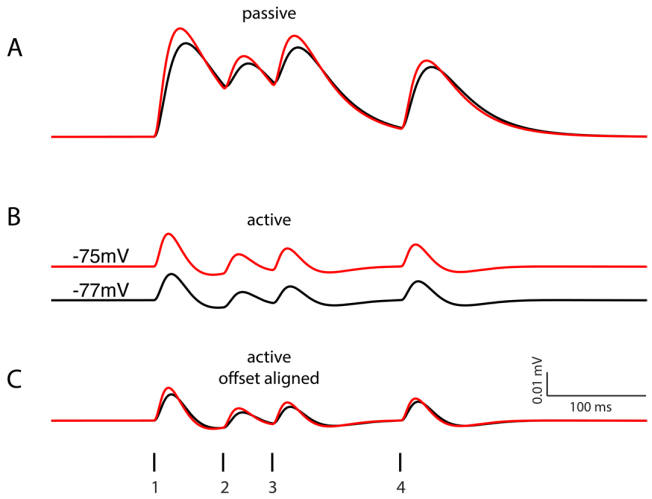

Figure S1. Voltage response for synaptic activation for passive and active models of L5 pyramidal cells. A. Sequential activation of the four synapses shown in **Figure 1C** in a passive model of L5 pyramidal cell (black) and its reduced model (red), as in **Figure 1E**. **B.** Activation of the same synapses but on the full active models of this cell (similar colors). Note that, in the active model, the resting potential in the reduced model is more depolarized than in the full model. **C.** same as in **B**, but when super positioning the resting potential in the two models.

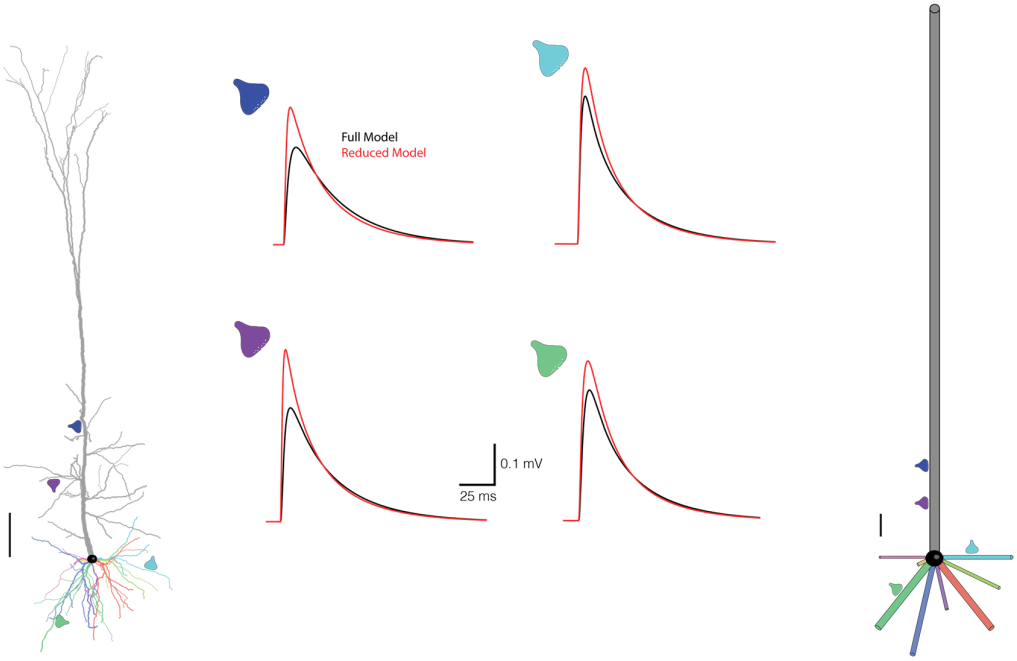

Figure S2. Somatic voltage response in full versus reduced model of layer 5 pyramidal cell. A. Left and right, the full and reduced models as in **Figure 1**. Middle, EPSP's measured at the soma of the full (black trances) and reduced (red trances) for the four simulated synapses. Location of synapses is shown by respective colored symbols on the tree and their EPSP by the respective colored synapses. Note that, as expected from **Figure 1D**, at high frequencies (e.g., at the EPSP peak) the reduced model overestimated the peak value at the soma, but the time-integral of the EPSP is more similar between the two models (they are identical, in the passive case, as the time-integral behave as does the steady-state case, see Rinzel and Rall, 1974). Synapse rise and decay time were 0.3 and 1.8 ms, recpectivly and the peak synaptic conductance was 0.8 nS. Scale bar at left is 100 μm in both full and reduced models.

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

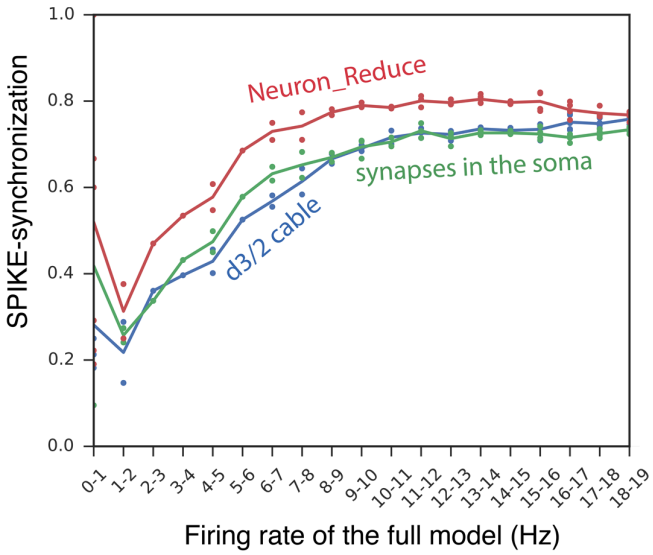

Figure S3: **The accuracy of *Neuron_Reduce* compared to two other simplification methods.** The accuracy is measured with SPIKE-synchronization as a function of the firing rate of the full model. Three reduction methods are compared: *Neuron_Reduce* (red), d3/2 reduction method of Rall ("equivalent cable" approach, blue), and mapping all the synapses to the soma (green), see **Methods** for details about the various reduction methods. Note that the accuracy of *Neuron_Reduce* is larger than in the other two methods over the entire range of model responses.

Reduction steps	Run time (s)	Run time speed-up factor
Full model	9,578 ± 3,500	1
Morphological reduction	6,917 ± 2,415	1.48 ± 0.46
Synaptic merging	2,173 ± 1150	5.00 ± 2.32
Neuron_Reduce	129 ± 50	79.47 ± 27.75

Table S1. Enhanced simulation speed in the reduced models versus the full models of L5 pyramidal cell. The modelled cell is shown in **Figure 1**; simulations were performed with 10,000 synapses for 50 seconds simulation time, excitatory and inhibitory synapses firing rate were 5 and 10 Hz, respectively (see **Figures 2** and **3**). Run time for the full model is shown in the first row. In the second row, the dendritic morphology was mapped to the respective multipolar cylindrical representation (as in **Figure 1B**) but the total number of synaptic point processes was kept as in the full model (1.48 folds reduction in run-time). In the third row, the full morphology remained untouched but synapses located on the same compartment were merged into a single point process (5 folds reduction in run-time). In the bottom row, the full *Neuron_Reduce* algorithm was used, whereby both the morphology reduction and the synaptic merging algorithms were implemented, resulting with ~80-folds reduction in simulation run-time (see also **Figure 3A**).

model	Number of synapses	AMPA strength (nS)	NMDA strength (nS)	Inhibitory strength (nS)	AMPA decay time constant (ms)*	NMDA decay time constant (ms)**	GABA decay time constant (ms)***	Excitatory activation rates (Hz)	Temp (°C)	Used in Figures	Citations and links
L5PC model	10000	0.4	0.28	1	1.7	43	8	3 - 8	37	2-5	(Hay et al., 2011) modelDB accession number 139653
L5PC no NMDA	10000	0.4	-	1	1.7	-	8	8 - 20	37	2	(Hay et al., 2011) modelDB accession number 139653
L5PC passive dends	10000	0.4	0.28	1	1.7	43	8	3 - 8	37	2	(Hay et al., 2011) modelDB accession number 139653
L5PC passive dends no NMDA	10000	0.4	-	1	1.7	-	8	8 - 20	37	1,2	(Hay et al., 2011) modelDB accession number 139653
Tufted PC (L6)	10000	0.79	0.56	0.88	1.74	43	7.65	1 - 3	34	6	(Markram et al., 2015) https://bbp.epfl.ch/nmc-portal/documents/10184/1921826/L6_TPC_L1_cADpyr231_5.zip
Large Basket Cell (L2/3)	1250	0.69	0.49	0.125	2	-	8	1 - 4	34	6	(Amsalem et al., 2016)
Double Bouquet Cell (L4)	2000	0.41	0.29	0.84	1.73	43	8.32	2 - 4	34	6	(Markram et al., 2015) https://bbp.epfl.ch/nmc-portal/documents/10184/52145/L4_DBC_cNAC187_1.zip
Spiny Stellate Cell (L4)	5000	0.76	0.54	0.84	1.74	43	7.82	1 - 4	34	6	(Markram et al., 2015) https://bbp.epfl.ch/nmc-portal/documents/10184/1921834/L4_SS_cADpyr230_1.zip
Martinotti Cell (L5)	5000	0.12	0.085	0.84	1.74	43	8.34	7 - 19	34	6	(Markram et al., 2015) https://bbp.epfl.ch/nmc-portal/documents/10184/52298/L5_MC_bAC217_2.zip
Spiny Rbp4-Negative (L4)	5000	0.4	0.28	1	1.7	43	8	3 - 5	34	6	(Gouwens et al., 2018) http://celltypes.brain-map.org/neuronal_model/download/483108201
Aspiny - Htr3a-Positive (L1)	5000	0.4	-	1	1.7	-	8	6 - 10	34	6	(Gouwens et al., 2018) http://celltypes.brain-map.org/neuronal_model/download/478045081
Human Pyramidal Cell (L2/3)	10000	0.7	-	0.7	1.8	-	8	2 - 3	37	6	(Eyal et al., 2016) https://senselab.med.yale.edu/ModelDB/showmodel.cshtml?model=195667

Table S2. Neuron models and synaptic parameters used in this paper

* the rise time constant for the AMPA synapses was 0.2 ms in all models besides the Large Basket Cell - 0.3 ms and the Human pyramidal cell - 0.3 ms

** the rise time constant for the NMDA synapses was 0.29 ms in all models with NMDA synapses

**** the rise time constant for the GABA synapses was 0.2 ms in all models besides the Large Basket Cell - 1 ms

bioRxiv preprint doi: https://doi.org/10.1101/506485; this version posted December 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.