bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Synaptic Dynamics as Convolutional Units

Julian Rossbroich[1◕], Daniel Trotter[2◕], Katalin Tóth[3], Richard Naud[2,3*]

1 Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
2 Department of Physics, University of Ottawa, Ottawa, ON, Canada
3 uOttawa Brain Mind Institute, Center of Neural Dynamics, Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada

◕These authors contributed equally to this work.
* rnaud@uottawa.ca

Abstract

Synaptic dynamics differ markedly across connections and strongly regulate how action potentials are being communicated. To model the range of synaptic dynamics observed in experiments, we develop a flexible mathematical framework based on a linear-nonlinear operation. This model can capture various experimentally observed features of synaptic dynamics and different types of heteroskedasticity. Despite its conceptual simplicity, we show it is more adaptable than previous models. Combined with a standard maximum likelihood approach, synaptic dynamics can be accurately and efficiently characterized using naturalistic stimulation patterns. These results make explicit that synaptic processing bears algorithmic similarities with information processing in convolutional neural networks.

Author summary

Understanding how information is transmitted relies heavily on knowledge of the underlying regulatory synaptic dynamics. Existing computational models for capturing such dynamics are often either very complex or too restrictive. As a result, effectively capturing the different types of dynamics observed experimentally remains a challenging problem. Here, we propose a mathematically flexible linear-nonlinear model that is capable of efficiently characterizing synaptic dynamics. We demonstrate the ability of this model to capture different features of experimentally observed data.

Introduction

The nervous system has evolved a communication system largely based on temporal sequences of action potentials. A central feature of this communication is that action potentials are communicated with variable efficacy on short (10 ms - 10 s) time scales [1–6]. The dynamics of synaptic efficacy at short time scales, or short-term plasticity (STP), can be a powerful determinant of the flow of information, allowing the same axon to communicate independent messages to different post-synaptic targets [7, 8]. Properties of STP vary markedly across projections [9–11], leading to the idea that connections can be conceived as belonging to distinct classes [12, 13] and that these distinct classes shape information transmission in vivo [14–16]. Thus, to

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

understand the flow of information in neuronal networks, the connectome must be indexed with an accurate description of STP properties.

One approach to characterizing synaptic dynamics is to perform targeted experiments and extract a summary feature, most commonly the paired-pulse ratio [5, 17–19], whereby a synapse can be classified as short-term depressing (STD) or short-term facilitating (STF). However, a single summary feature is insufficient to capture the full extent of STP diversity. Longer or more complex stimulation patterns are required to describe delayed facilitation onset [6], biphasic STP [20, 21] or the distinction between supra- and sub-linear facilitation [22]. Such atypical STP dynamics challenge the traditional dichotomy of STF and STD and suggest that more complex phenotypes can exist and contribute to network function in unknown ways. A complimentary approach to characterizing STP is to fit a mechanistic mathematical model using all available experimental data, where the parameters correspond to physical properties [23]. In this vein, the model proposed by Tsodyks and Markram captures the antagonism between transient increases in vesicle release probability and transient depletion of the readily releasable vesicle pool [11, 24, 25]. Optimizing the parameter values to best fit the observed data provides an estimate of biophysical properties, and thus supports a classification of STF or STD [26, 27]. This simplicity makes the Tsodyks-Markram model highly interpretable, but restricts its ability to capture the diversity of synaptic responses to complex stimulation patterns. Complex STP dynamics rely on interactions between multiple synaptic mechanisms, that cannot be described in a simplified framework of release probability and depletion. To describe the dynamics of complex synapses, the Tsodyks-Markram model therefore requires multiple extensions [23, 28], for instance, vesicle priming, calcium receptor localization, multiple timescales, or use-dependent replenishment [6, 29–31]. As a compendium of biophysical properties is collected, the properties become increasingly difficult to adequately characterize based on experimental data because degeneracies and over-parametrization lead to inefficient and non-unique characterization. Taken together, current approaches appear to be either too complex for accurate characterization, or insufficient to capture all experimental data.

The trade-off between the model's identifiability and its ability to espouse complex experimental data echoes similar trade-offs in other fields, such as in the characterization of the input-output function of neurons [32–37]. Taking a systems identification approach, we seek to sacrifice some of the model's interpretability in order to avoid over-parametrization and degeneracies while still capturing the large range of synaptic capabilities. Inspired by the success of linear-nonlinear models for the characterization of cellular responses [32, 33], we extend previous phenomenological approaches to synaptic response properties [3, 4, 38, 39] to take into account nonlinearities and the kinetics evolving on multiple time scales. We show that the resulting Spike Response Plasticity (SRP) model captures short-term facilitation, short-term depression, biphasic plasticity, as well as supralinear facilitation and post-burst potentiation. Using standard gradient descent algorithms, model parameters can be inferred accurately with limited amount of experimental data. Our work also makes explicit that synaptic dynamics extend the information processing of dendritic integration by adding another layer of convolution combined with nonlinear readout as in deep convolutional neural networks, thus contributing to build a theory for information processing at synapses.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

Deterministic Dynamics

To construct our statistical framework, we first consider the deterministic dynamics of synaptic transmission. Our goal is to describe the dynamics of the amplitude of individual post-synaptic currents (PSCs). Specifically, a presynaptic spike train will give rise to a post-synaptic current trace, $I(t)$, made of a sum of PSCs triggered by presynaptic action potentials at times t_j:

$$I(t) = \sum_j \mu_j k_{PSC}(t - t_j), \tag{1}$$

where k_{PSC} is the stereotypical PSC time course and μ_j is the synaptic efficacy, or relative amplitude, of the jth spike in the train normalized to the first spike in the train ($\mu_1 = 1$).

To begin modeling synaptic dynamics, we seek a compact description for generating $I(t)$ from the presynaptic spike train, $S(t)$. Spike trains are mathematically described by a sum of Dirac delta-functions, $S(t) = \sum_j \delta(t - t_j)$ [35]. For our present purposes, the time course of individual PSCs is assumed to remain invariant through the train, but PSC amplitude is dynamic. To capture these amplitude changes, we introduce the concept of an efficacy train, $E(t)$, made of a weighted sum of Dirac delta-functions: $E(t) = \sum_j \mu_j \delta(t - t_j)$. The efficacy train can be conceived as a multiplication between the spike train and a time-dependent signal, $\mu(t)$, setting the synaptic efficacy at each moment of time

$$E(t) = \mu(t)S(t). \tag{2}$$

Thus the current trace can be written as a convolution of the efficacy train and the stereotypical PSC shape, $\mathbf{k}_{PSC}$: $I = \mathbf{k}_{PSC} * E$, where $*$ denotes a convolution. In this way, because in typical electrophysiological assays of synaptic properties the PSC shape ($\mathbf{k}_{PSC}$) is known and the input spike train $S(t)$ is controlled, characterization of synaptic dynamics boils down to a characterization of how the synaptic efficacies evolve in response to presynaptic spikes. Mathematically, we seek to identify the functional $\mu[S(t)]$ of the spike train $S(t)$.

Using this formalism, we aim to build a general framework for capturing synaptic efficacy dynamics. Previous modeling approaches of STP have used a system of nonlinear ordinary differential equations to capture $\mu(t)$ separated in a number of dynamic factors [4, 11, 23, 24]. Our main result is that we propose a linear-nonlinear approach inspired from the engineering of systems identification [33, 40–47] and the Spike Response Model (SRM) for cellular dynamics [34, 48, 49]. Here, the efficacies are modeled as a nonlinear readout, f, of a linear filtering operation:

$$\mu = \frac{1}{f(b)} f\left(\mathbf{k}_\mu * S + b\right) \tag{3}$$

where $\mathbf{k}_\mu(t)$ is the *efficacy kernel* describing the spike-triggered change in synaptic efficacy and b is a baseline parameter, which could be absorbed in the definition of the efficacy kernel. The efficacy kernel can be parametrized by a linear combination of nonlinear basis functions (see Methods). In this way, while $\mathbf{k}_{PSC}$ regulates the stereotypical time-course of a single PSC, the efficacy kernel, $\mathbf{k}_\mu$, regulates the stereotypical changes in synaptic efficacy following a pre-synaptic action potential. The efficacy kernel can take any strictly causal form ($k_\mu(t) = 0$ for $t \in -\infty, 0]$), such that a spike at time t_j affects neither the efficacy before t_j nor the efficacy at time t_j. Here we call the 'potential efficacy' the result of the convolution and baseline, $\mathbf{k}_\mu * S + b$, before taking a sigmoidal nonlinear readout. Although some early studies have used a linear

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

readout [4], a nonlinear readout is potentially more apt to capture both the frequency dependence and the nonlinear progression of PSC amplitudes in response to periodic stimulation. The factor $f(b)^{-1}$ is introduced because we are considering amplitudes normalized to the first pulse, but can be replaced by an additional parameter when treating non-normalized amplitudes. Together, the deterministic SRP model presented so far, could capture different types of STP by changing the shape of the efficacy kernel.

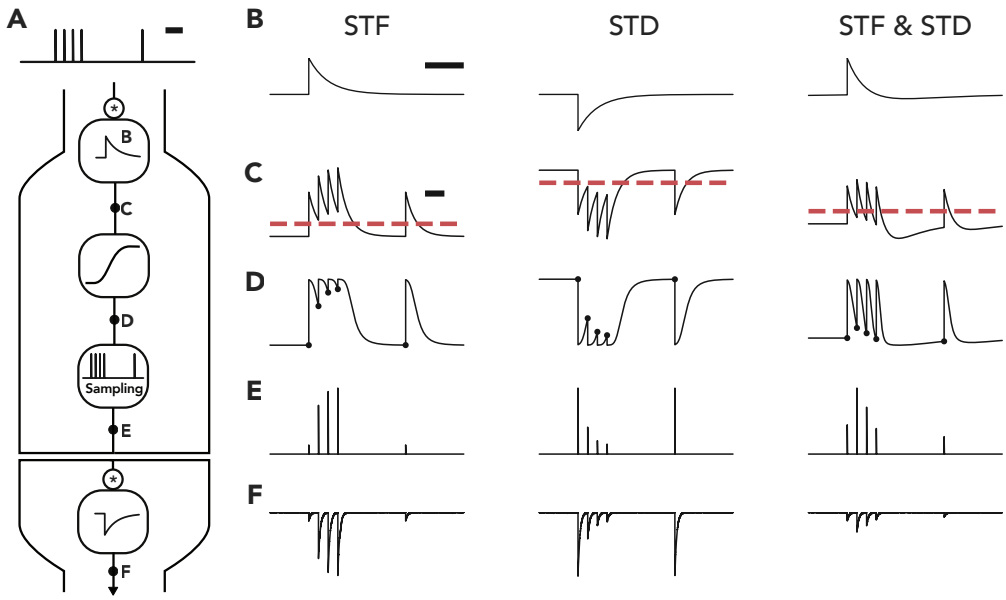

Fig 1. The SRP model captures different types of short-term plasticity. **A** The model first passes a pre-synaptic spike train through a convolution with the impulse-response change in efficacy. We illustrate three choices of this efficacy kernel (**B**), a positive kernel for STF (left), a negative kernel for STD (middle) and one for STF followed by STD (right). After the convolution and combination with a baseline (**C**; dashed line indicates zero), a nonlinear readout is applied, leading to the time-dependent efficacy $\mu(t)$ (**D**). This time-dependent signal is then sampled at the spike times, leading to the efficacy train (**E**) and thus to the post-synaptic current trace (**F**). Scale bars correspond to 100 ms.

Short-Term Facilitation and Depression

To show that the essential phenomenology of both STF and STD can be encapsulated by an efficacy kernel $\mathbf{k}_\mu$, we will study the response to a burst of four action potentials followed with a delay by a single spike and compare the responses obtained when changing the shape of the efficacy kernel (Fig. 1A). For simplicity, we consider $\mathbf{k}_\mu$ to be a mono-exponential decay starting at time 0. When the amplitude of this filter is positive (Fig. 1B, left), a succession of spikes will lead to an accumulation of potential efficacy ($\mathbf{k}_\mu * S + b$, Fig. 1C, left). After the sigmoidal readout (Fig. 1D, left) and sampling at the spike times, the efficacy train (Fig. 1E, left) and the associated current trace (Fig. 1F, left) show facilitation. Choosing a negative amplitude (Fig. 1B, middle) gives rise to the opposite phenomenon. In this case, the succession of spikes will gradually decrease potential efficacy ($\mathbf{k}_\mu * S + b$, Fig. 1C, middle). Following the sigmoidal readout (Fig. 1D, middle) the efficacy train (Fig. 1E, middle) and the

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

resulting current trace (Fig. 1F, middle) show STD dynamics. Conveniently, changing the polarity of the efficacy kernel controls whether synaptic dynamics follow STF or STD.

At many synapses, facilitation apparent at the onset of a stimulus train is followed by depression, a phenomenon referred to as biphasic plasticity [20, 21, 50]. To model this biphasic plasticity in our framework, we consider an efficacy kernel made of a combination of two exponential-decays having different decay timescales and opposing polarities. By choosing the fast component to have a positive amplitude and the slow component to have a negative amplitude (Fig. 1B, right), we obtain a mixture between the kernel for STF and the kernel for STD. Then, a succession of spikes will create an accumulation of potential efficacy followed by a depreciation ($\mathbf{k}_\mu * S + b$, Fig. 1C, right). Once the sigmoidal readout is performed (Fig. 1D, right), the efficacy train (Fig. 1E, right) and the resulting PSC trace (Fig. 1F, right) show facilitation followed by depression. Thus, the model captures various types of STP by reflecting the facilitation and depression in positive and negative components of the efficacy kernel, respectively.

Sublinear and Supralinear Facilitation

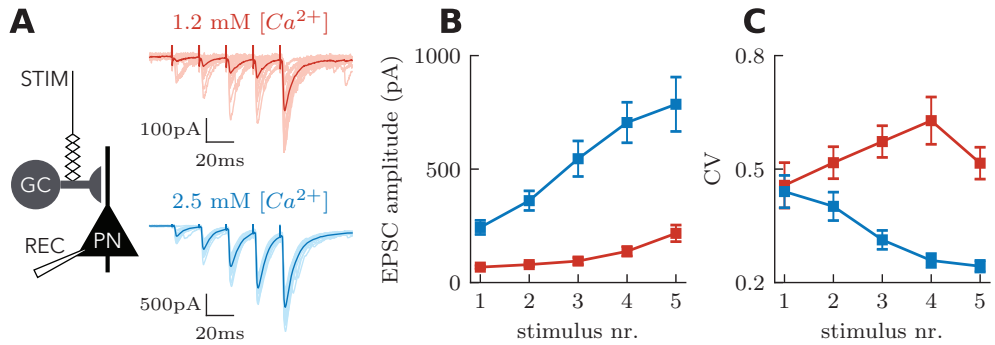

Fig 2. Effects of extracellular calcium concentration on STP dynamics at hippocampal mossy fiber synapses. **A** Mossy fiber short-term facilitation in 1.2 mM (red) and 2.5 mM (blue) extracellular [Ca^{2+}]. EPSCs recorded from CA3 pyramidal cells in response to stimulation of presynaptic mossy fibers (50 Hz, 5 stimuli). **B** EPSC peak amplitudes as a function of stimulus number. The time course of facilitation varies dependent on the initial release probability. **C** The coefficient of variation (CV), measured as the standard deviation of EPSCs divided by the mean, is increased in 1.2 mM extracellular [Ca^{2+}]. Data redrawn from [22].

The typical patterns of facilitation and depression shown in Fig. 1 are well captured by the traditional Tsodyks-Markram (TM) model [24–26]. We, therefore, ask if our modeling framework can capture experimentally observed features that require a modification of the classical TM model. While previous work has extended the TM model for use-dependent depression [29] and receptor desensitization [23], we consider the nonlinear facilitation observed in mossy fiber synapses onto pyramidal neurons (MF-PN) in response to a burst of action potentials (Fig. 2A). In these experiments, the increase of PSC amplitudes during the high-frequency stimulation is nonlinear. Interestingly, the facilitation is sublinear at normal calcium concentrations (2.5 mM extracellular [Ca^{2+}]), but supralinear in physiological calcium concentrations (1.2 mM extracellular [Ca^{2+}]) [22] (Fig. 2B). The supralinearity of STF observed in 1.2 mM [Ca^{2+}] is caused by a switch from predominantly univesicular to predominantly multivesicular release. In contrast, multivesicular release is already in place in 2.5 mM [Ca^{2+}], and the facilitation observed under these conditions can be solely attributed to

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the recruitment of additional neurotransmitter release sites at the same synaptic bouton [22]. These two mechanisms, by which MF-PN synapses can facilitate glutamate release, are generated by complex intra-bouton calcium dynamics [30, 51, 52], which lead to gradual and compartmentalized increases in calcium concentration. Consistent with the expectation that these two modes could lie on the opposite sides of the inverse-parabolic relationship between coefficient of variation (CV) and mean, normal calcium is associated with a gradual decrease of CV through stimulation, while physiological calcium is associated with an increase of CV (Fig. 2C). Perhaps because the TM model was based on experiments at 2 mM calcium concentration, the model emulates sublinear facilitation. Supralinear facilitation is not possible in the original structure of the model (Fig. 3C), as can be verified by mathematical inspection of the update equations (see Methods). Hence the TM model must be modified to capture the supralinear facilitation typical of experimental data at physiological calcium concentrations.

To extend the TM model to account for supralinear facilitation, we considered a small modification to the dynamics of facilitation without adding a new dynamic variable (Fig. 3A), although supralinear facilitation can be achieved with an additional state variable. This modification allows the facilitation variable of the TM model to increase supralinearly when this variable is small, and sublinearly when this variable is large (see Methods). By lowering the baseline facilitation parameter, the extended TM model switches from sublinear facilitation to a supralinear facilitation (Fig. 3D). We thus have shown that a modification to the set of equations for the TM model is required to present supralinear facilitation and capture the experimentally observed facilitation at physiological calcium.

In contrast, for the linear-nonlinear model framework, the switch from sublinear to supralinear facilitation does not require a modification to the equations. We can change sublinear facilitation into a supralinear one by lowering the baseline parameter without changing the efficacy kernel. When the baseline parameter is high, a facilitating efficacy kernel is likely to hit the saturating, sublinear, part of the nonlinear readout (Fig. 3E). When the baseline parameter is low, the same facilitating efficacy kernel can recruit the onset of the nonlinearity, which gives rise to supralinear facilitation (Fig. 3F). Thus, the changes in extracellular calcium are conveniently mirrored in a modification of a baseline parameter in the SRP model.

Facilitation Latency

Next we illustrate the role of the efficacy kernel to generalize to the multiple timescales of STP without requiring a change in the structure of the model. As an illustrative example, we will focus on the one particular synapse showing facilitation latency [6]. In mossy fiber synapses onto inhibitory interneurons, the facilitation caused by a burst of action potentials increases during the first 2 seconds after the burst (Fig. 4A). This delayed facilitation cannot be captured by the classical TM model because facilitation is modeled as a strictly decaying process and the experimental data shows that facilitation increases during the first 1-2 seconds following the burst. Adding to this model a differential equation for the slow increase of facilitation is likely sufficient to capture facilitation latency, but this modification consists of a significant modification to the modeling framework. In the linear-nonlinear framework, one could capture the facilitation latency by modifying the shape of the efficacy kernel. An efficacy kernel with a slow upswing (Fig. 4B), once convolved with a burst of action potential followed by a test-pulse (Fig. 4C) will produce a delayed increase in synaptic efficacy (Fig. 4D) as well as match the nonlinear increase in facilitation with the number of stimulation spikes. Without automated fitting of the kernel to the data, a simple change to the efficacy kernel captures facilitation latency. Thus, provided that the efficacy kernel is

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

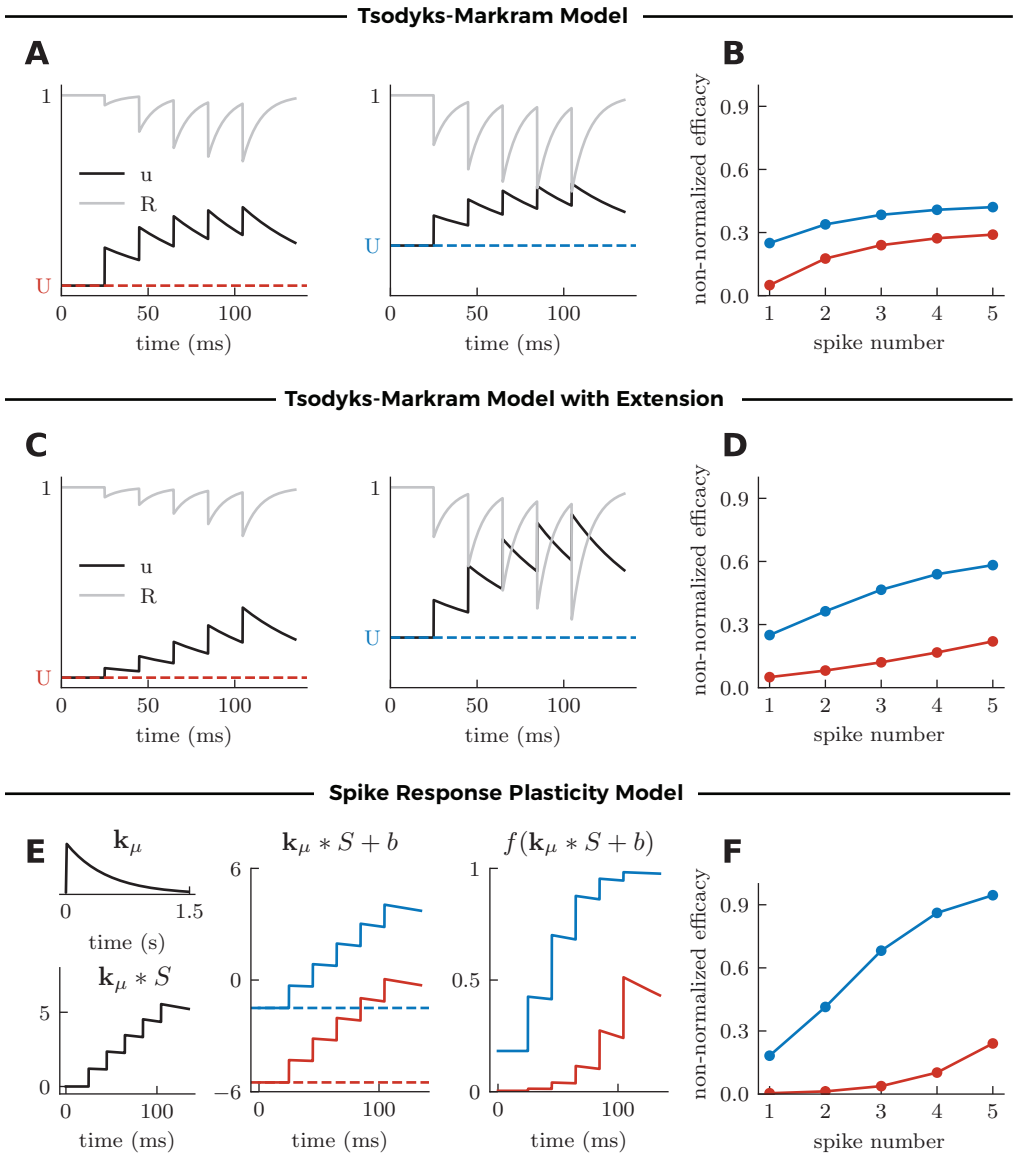

Fig 3. Modeling sublinear and supralinear facilitation through changes in the baseline parameter. **A** Mechanism of the classic TM model [24–26], illustrated in response to 5 spikes at 50 Hz for different values of the baseline parameter U. **B** Synaptic efficacy $u * R$ at each spike according to the classic TM model. Facilitation is always restricted to sublinear dynamics. **C** Mechanism and **D** Synaptic efficacy $u * R$ at each spike according to the extended TM model (see Methods). Choosing the baseline parameter U sufficiently small allows for supralinear facilitation. **E** Mechanism of the SRP model, illustrated for two different values of the baseline parameter b, with the same synaptic efficacy kernel k_μ (left). Changing the baseline parameter b leads to a linear displacement of the filtered spike train $k_\mu * S + b$ (middle), which causes a shift from sub- to supralinear dynamics after the nonlinear readout $f(k_\mu * S + b)$. **F** Resulting synaptic efficacy at each spike according to the SRP model. Changing the baseline parameter causes a switch from sublinear to supralinear facilitation, as observed experimentally in response to varying extracellular $[Ca^{2+}]$ (see Fig 2.)

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

parameterized with basis function spanning a large part of the function space, the SRP model can aptly generalize to STP properties unfolding on multiple timescales.

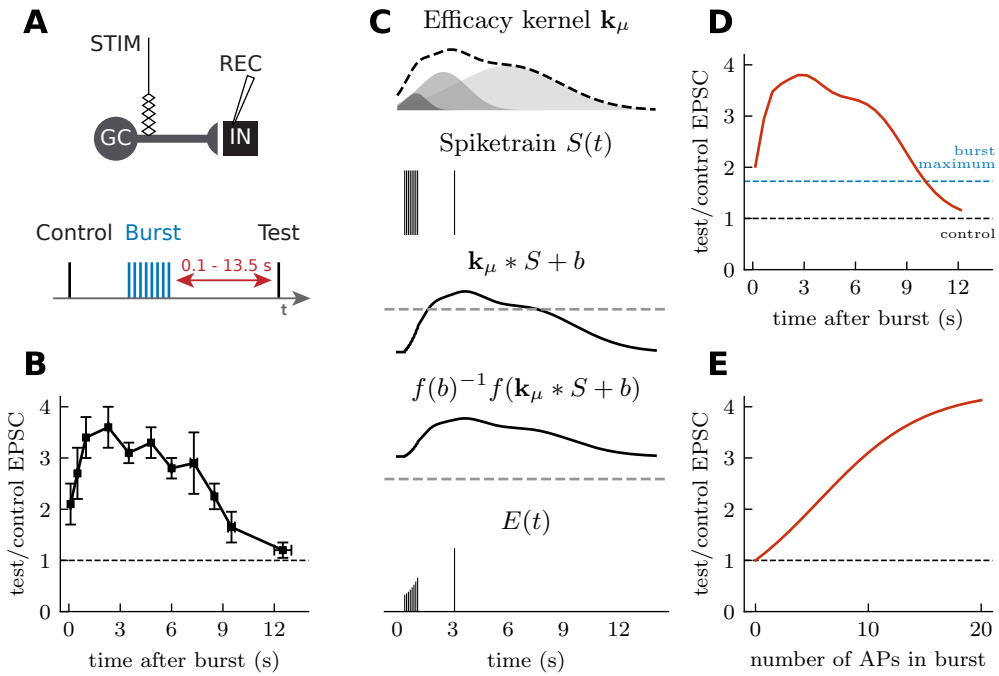

Fig 4. Post-burst facilitation captured by a delayed facilitation kernel. **A** Experimental setup and **B** measurement of post-burst facilitation in CA3 interneurons (redrawn from Ref. [6]). **C** Synaptic plasticity model. A delayed facilitation kernel was chosen as a sum of three normalized Gaussians with amplitudes {450, 2200, 5500}, means {1.0, 2.5, 6.0} s and standard deviation {0.6, 1.3, 2.8} s. The spiketrain (8 spikes at 100 Hz followed by a test spike) is convolved with the delayed facilitation kernel. A nonlinear (sigmoidal) readout of the filtered spike train leads synaptic efficacies. Dashed lines indicate zero. **D** Efficacies of test spikes in the synaptic plasticity model as a function of the number of action potentials in the preceding burst. **E** Synaptic efficacy of test spikes (3 s after a single burst at 160 Hz) as a function of the number of action potentials (APs).

Stochastic Properties

Synaptic transmission is inherently probabilistic. The variability associated with synaptic release depends intricately on stimulation history, creating a complex heteroskedasticity. Figure 2C illustrates one type of heteroskedasticity observed experimentally, whereby the variability increases through a stimulation train but only for the physiological calcium condition. To capture these transmission properties, we must establish a stochastic framework.

In the previous section, we have treated the deterministic case, which corresponds to the average synaptic efficacies. We now consider a sample of synaptic efficacies to be a random variable such that the jth spikes associated with the random variable Y_j. Its mean is given by the linear-nonlinear operation:

$$\langle Y_j \rangle \equiv \mu_j = \frac{1}{f(b)} f\left(\mathbf{k}_\mu * S(t_j) + b\right). \tag{4}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

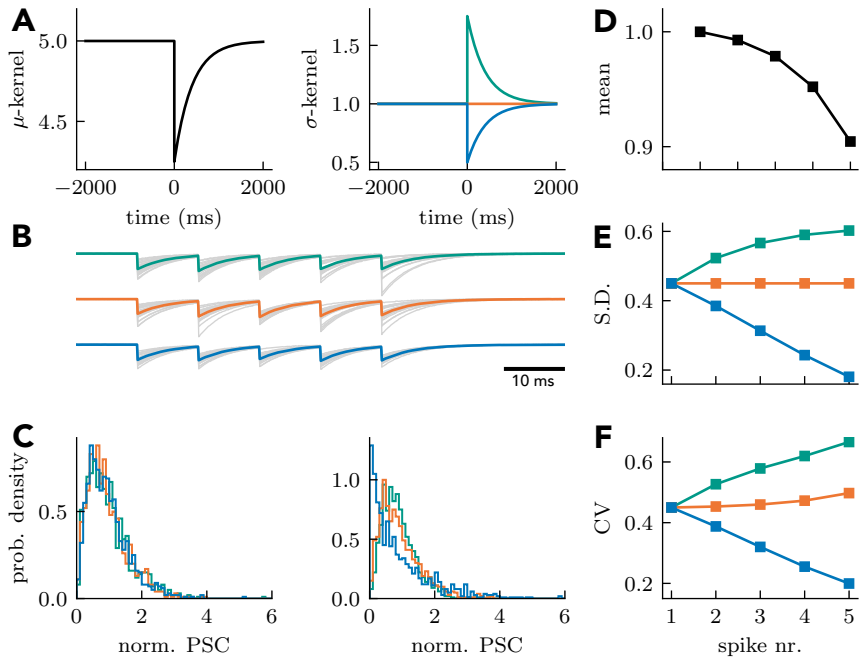

Fig 5. Capturing heteroskedasticity with a two-kernel approach. **A** The μ-kernel regulating the dynamics of the mean amplitude is paired with a σ-kernel regulating the dynamics of the variance. Three σ-kernels are shown: a variance increasing (teal), a variance invariant (orange) and a variance decreasing (blue) kernel. **B** Sample PSC responses to a spike train generated from the three σ-kernels (gray lines) along with the associated mean (full lines). **C** Probability density function of the amplitude of the first (left) and last (right) pulse. **D** The mean amplitude is unaffected by different σ-kernels. **E** The standard deviation is either increasing (teal), invariant (orange) or decreasing (blue), consistent with the polarity of the σ-kernel. **F** The coefficient of variation results from a combination of μ and σ kernel properties.

In this way, the current trace is made of PSCs of randomly chosen amplitudes whose average pattern is set by the efficacy kernel: $I(t) = \sum_j y_j \mathbf{k}_{PSC}(t - t_j)$, where y_j is an instance of Y_j. Sampling from the model repeatedly will produce slightly different current traces, as is typical of repeated experimental recordings (Fig. 3A).

To establish the stochastic properties, we must choose a probability distribution for the synaptic efficacies. Previous work has argued that the quantal release of synaptic vesicles produces a binomial mixture of Gaussian distributions [53, 54]. There is substantial evidence, however, that releases at *single synapses* are better captured by a mixture of skewed distribution such as the binomial mixture of gamma distributions [55, 56]. Such skewed distributions are also a natural consequence of Gaussian-distributed diameters and the cubic transform of vesicle volumes [57]. For multiple synaptic contacts, release amplitudes should then be captured by a weighted sum of such binomial mixtures, a mixture of mixtures as it were. Indeed, a binomial mixture of skewed distributions has been able to capture the stochastic properties of PSC amplitudes from multiple synaptic contacts [27, 58], but only under the assumption that each synapse contributes equally to the compound PSC. Together, these considerations suggest that in seeking a simple parameterization of the random process, we must find a skewed distribution whose mean and standard deviation can change in the course of STP.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Following prior work [56, 58], we chose to focus on gamma-distributed PSCs:

$$p(y_j|S(t_j), \theta) = g(y_j|\mu_j, \sigma_j), \tag{5}$$

where $g(y|\mu, \sigma)$ is the gamma distribution with mean, μ, and standard deviation, σ. The mean is set by the linear-nonlinear operation (Eq. 4) and the standard deviation is set by a possibly distinct linear-nonlinear operation:

$$\sigma_i = \sigma_0 f(k_\sigma * S_i + b_\sigma), \tag{6}$$

where we have introduced a baseline parameter, b_σ and another kernel, k_σ, for controlling the standard deviation. We call this time-dependent function, the variance kernel. The factor σ_0 is introduced to scale the nonlinearity f appropriately, but could be omitted if data has been standardized. In this framework, some common statistics have a simple expression in terms model parameters. This is the case for the stationary CV. Since we are considering filters decaying to zero after a long interval, the statistics of releases arriving after a long interval depends solely on the baseline parameters, such that CV $= \sqrt{1/\sigma_0 f(b_\sigma)}$.

The properties of this choice of probability distributions are illustrated in Figure 5. Using a depressing kernel, Fig. 5 depicts the effect of choosing a variance kernel with positive, negative and zero amplitude (Fig. 5A). These kernel choices show that the model can capture both increases and decreases of variability, although an increase in variability during STD is generally observed [59, 60]. The temporal profile of the variance kernel determines the time-dependent changes in variance. For simplicity, we chose an exponential decay with a relaxation time scale equal to that of the efficacy kernel. The kernel amplitude and baseline were chosen to match experimental observations at STD synapses (CV increasing from a little less than 0.5 to almost 1 after 5 pulses [60]). With these modeling choices, we simulated the probabilistic response to input trains (Fig. 5B, 5 spikes, 100 Hz). The model with positive σ-kernel shows a progressive increase of trial-to-trial variability. Conversely, the model with a negative σ-kernel displays the opposite progression, as can be observed by comparing the probability distribution of the first and the last response (Fig. 5C). The average response follows precisely the same STD progression (Fig. 5D), despite drastically different progression of standard deviation (Fig. 5E) and CV (Fig. 5F). Thus gamma-distributed amplitudes with dynamic variance can capture multiple types of heteroskedasticity.

Next we asked if the model could capture the striking changes in heteroskedasticity observed in some experiments (Fig. 2C). In this case, decreasing the extracellular concentration of calcium not only changed the average response progression from sublinear to supralinear (Fig. 2B), but also changed the CV progression from strongly decreasing to strongly increasing (Fig. 2C, [22]). Figure 6 shows that changing the μ-kernel baseline in a model with facilitating standard deviation can reproduce this phenomenon. Here, as in the deterministic version of the model, the change in baseline changes the progression of efficacies from sublinear to supralinear (Fig. 6A-D). These effects are associated with changes in variances that are sublinear and supralinear, respectively (Fig. 6E). In the model with a low baseline (red curve in Fig. 6), the variance increases more quickly than the efficacy, leading to a gradual increase in CV. Despite the fact that the variance increases for both cases (Fig. 6E), only the model with sublinear increase in efficacy displays a decreasing CV. We conclude that, by controlling a baseline parameter, the model can capture both the change from sublinear to supralinear facilitation and the change in heteroskedasticity incurred by a modification of extracellular calcium.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

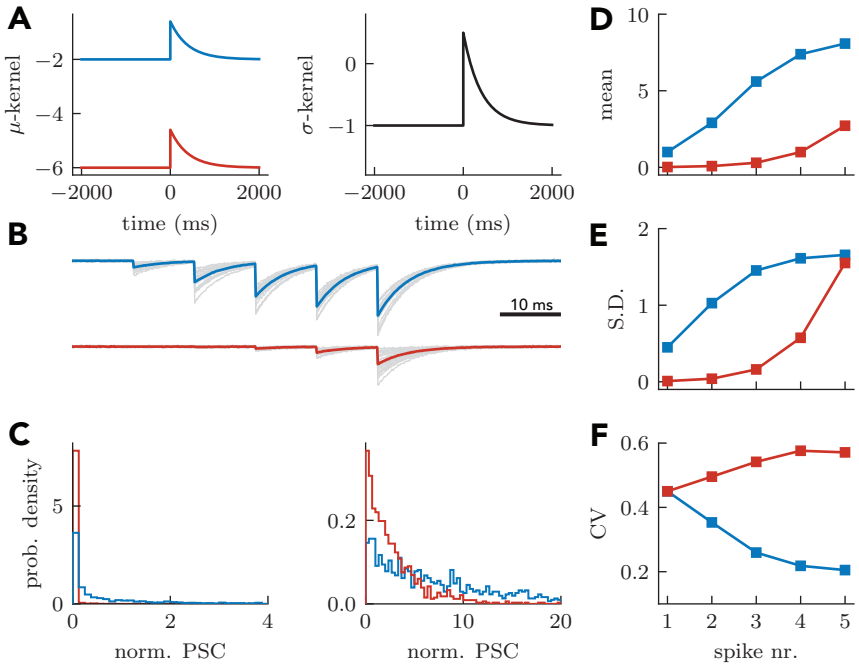

Fig 6. Capturing effect of external calcium concentration on coefficient of variation through baseline of μ-kernel. **A** Comparing facilitating μ-kernels with high (blue) and low (red) baseline but fixed, σ-kernel. **B-F** as in Fig. 5. The coefficient of variation increases with pulse number for the low baseline case, but decreases with pulse number for the high baseline case.

Inference

Thus far, we have illustrated the flexibility of the SRP framework for qualitatively reproducing a diversity of notable synaptic dynamics features. Now we investigate the ability of this framework to capture synaptic dynamics quantitatively. As in the characterization of cellular dynamics [61], a major impediment to precise characterization is parameter estimation. Since efficient parameter inference is largely depends on the presence of local minima, we first investigated the cost function landscape for estimating model parameters.

We have developed an automatic characterization methodology based on the principle of maximum likelihood (see Methods). Given our probabilistic model of synaptic release, we find optimal filter time-course by iteratively varying their shape to determine the one maximizing the likelihood of synaptic efficacy observations. The method offers a few advantages. Firstly, the method is firmly grounded in Bayesian statistics, allowing for the inclusion of prior knowledge and the calculation of posterior distribution over the model parameters [26, 58]. Secondly, although targeted experiments can improve inference efficiency, our approach does not rely on experimental protocols designed for characterization. Naturalistic spike trains recorded *in vivo* [30, 62], Poisson processes or other synthetic spike trains can be used in experiments to characterize synaptic dynamics in realistic conditions.

To test the efficiency of our inference method, we generated an artificial Poisson spike train and used this spike train to generate surrogate synaptic efficacy data using our SRP model (Fig. 7A-B). We then asked if our inference method identified the correct parameters and whether local minima were observed. Instead of the case where the filters are described by a combination of nonlinear basis functions, we considered

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

only one basis function, a mono-exponential decay with its decay time constant known. In cases where the time constant is unknown, one would fit the coefficient of a combination of nonlinear basis functions, as is typical in other linear-nonlinear models [32, 34, 63, 64]. Using a long stimulus train, the likelihood function appeared convex over a fairly large range of parameter values as no local minima were observed (Fig. 7C-F). The slanted elongation of likelihood contour indicates a correlation or anti-correlation between parameter estimates. Not surprisingly, we found that the estimates of baseline and scale factor of the σ-kernel would be anti-correlated (Fig. 7D), while on the other hand the estimates of filter amplitudes for efficacy and variance would have little correlation (Fig. 7C). Furthermore, we found that the parameter estimates matched closely the parameters used to simulate the responses after 100 to 150 spikes (Fig. 7G-H), with efficacy parameters requiring more data than variance parameters (compare Fig. 7G and H). The relationship between error in parameter estimation and training size is such that for large training sets the percent error goes to zero (Fig. 7G). As might be expected, the method performs poorly for parameters that do not regulate the efficacy. For instance, when a facilitating efficacy kernel is added to a high baseline parameter, the high baseline saturates the nonlinear readout, and no facilitation of the efficacy will be observed. As a result, kernel amplitude is poorly estimated when the baseline used for simulations is high (Fig. 7H). Using a separate artificial Poisson input for testing the predictive power of the model, we calculated the mean squared error between the inferred and true model (Fig. 7I). The prediction error of the inferred model almost matched that of the true model, even if inference was based on less than 100 spikes. We conclude that maximum likelihood applied to the SRP model is able to characterize the model efficiently and accurately, and that, for simple filters, the landscape is sufficiently devoid of local minima to allow efficient characterization.

Relation to Generalized Linear Models

We have shown that, in one situation, the likelihood landscape appears devoid of local minima, but is this always the case? Without additional restrictions on the model described in the previous section, it is unlikely that the likelihood would be always convex. However, with some simplifications, the model becomes a Generalized Linear Model (GLM), and convexity proofs are possible in some cases [41, 65]. In this section, we describe two such simplifications.

We can assume that the standard deviation is always proportional to the mean: $\boldsymbol{\sigma} = \sigma_0 \boldsymbol{\mu}$. This assumes that the CV is constant through a high-frequency train, a coarse assumption given the large changes in CV observed experimentally [22, 60]. If for some reason an accurate reproduction of the changes in variability can be sacrificed, this simplification leads to interesting properties. In this case, no variance parameters are to be estimated apart from the scaling σ_0. There is thus a reduction in the number of parameters to be estimated. In addition, since the gamma distribution belongs to the exponential family and the mean is a linear-nonlinear function of the other parameters, we satisfy the requirements for GLMs and the likelihood is devoid of local minima when the proof given by Paninski (2004) [41] applies.

For the depressing synapses, the CV is increasing during a high-frequency train. This can be modeled by a constant standard deviation with a mean decreasing through the stimulus train. Similarly, for the facilitating synapses at normal extracellular calcium shown in Fig. 2, the gradual decrease in CV can be explained by an approximately constant standard deviation, $\boldsymbol{\sigma} = \sigma_0$, and an increasing mean. Setting the variance to a constant again reduces the number of parameters to be estimated and recovers the necessary assumptions of a GLM.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

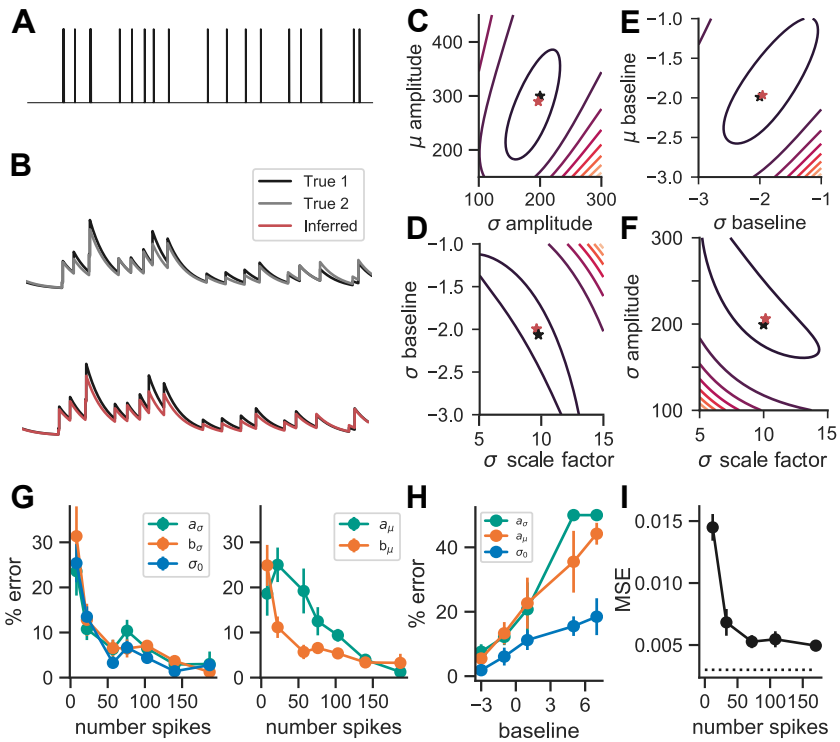

Fig 7. Statistical inference of kinetic properties on surrogate data. **A** Simulated Poisson spike trains mark pre-synaptic stimulation. **B** Simulated post-synaptic currents of the spike train in A for two independent sampling of the true parameter set (top) and for sampling of true and inferred models (bottom). Negative log-likelihood landscape as a function of **C** μ- and σ-kernel amplitudes, **D** σ baseline and scaling factor, **E** μ and σ baseline and **F** σ scaling and amplitude. **G** average σ parameter errors as function training size (left) and average μ parameter errors as a function of training size (right). **H** Average amplitude (μ and σ) and σ scaling factor error as a function of saturation; error bars are SEM. **I** Mean square error (MSE) of inferred and true model as a function of training size. Dashed line is MSE between independent samples of the true parameter set.

Relation to Convolutional Neural Networks

A convolution followed by a nonlinear readout is also the central operation performed in convolutional neural networks (CNNs). Because this type of algorithm is associated with high performance in challenging tasks, we asked what type of neural network architecture corresponds to synaptic information processing. In a artificial neural network, an input arranged as a one-dimensional array $\mathbf{x}$ is convolved with a bank of kernels $\{\mathbf{k}_i\}$ and readout through a nonlinearity f to generate a representation of a first hidden layer of neurons

$$h_t^{(i)} = f\left(\mathbf{k}_i^T(\mathbf{m} \odot \mathbf{x}_{t:t+K})\right) \tag{7}$$

where K_i is the length of the ith kernel and a mask $\mathbf{m}$ operates on the input with the Hadamard product ($\odot$). This mask is made of samples from Bernoulli random variable normed so that the average of $\mathbf{k}_i^T(\mathbf{m} \odot \mathbf{x}_{t:t+K})$ is $\mathbf{k}_i^T \odot \mathbf{x}_{t:t+K}$. It randomly silences

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

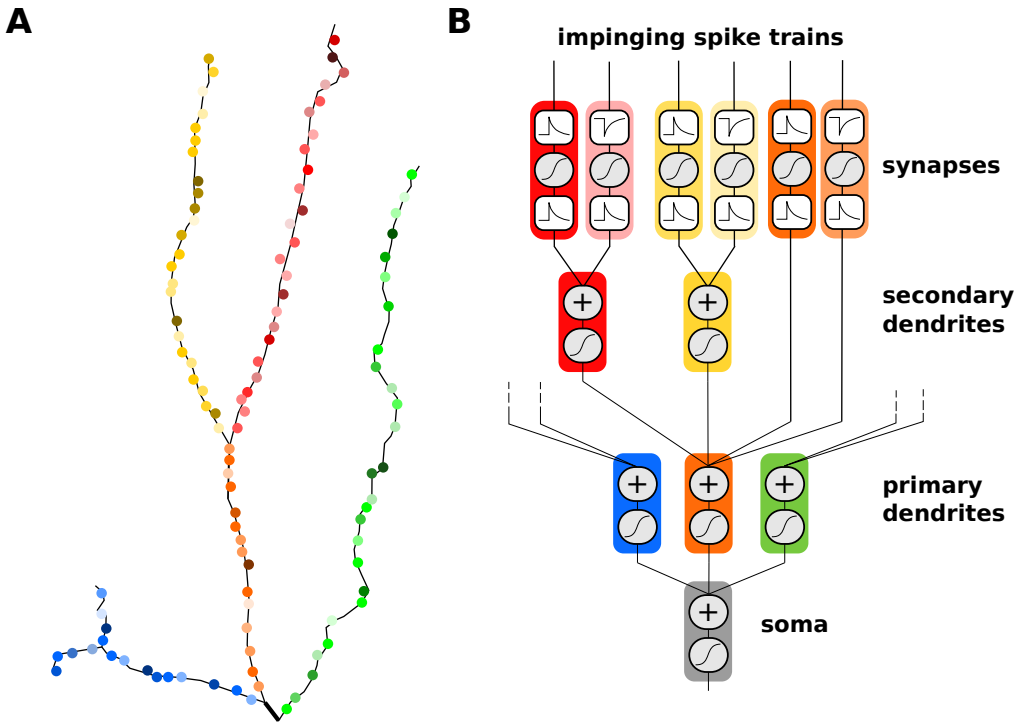

Fig 8. A synaptic contribution to the hierarchy of linear-nonlinear computations. **A** Synapses distributed on primary (orange, blue and green) and secondary (yellow and red) dendrites may have different synaptic properties (different color tints). **B** Each synapse is characterized by two kernels separated by a nonlinear sampling operation. (1) A pre-synaptic convolution kernel regulates synaptic dynamics. (2) A post-synaptic convolution kernel regulates the shape of the post-synaptic potential locally. The post-synaptic potentials from different synapses are summed within each dendritic compartment, forming a processing hierarchy converging to the soma.

parts of the input, an approach called *dropout* that was proposed to improve learning [66]. Although CNN architectures vary, the next layer may be that of a pooling operation $h_k = \frac{1}{Z}\sum_{t=k}^{k+Z} h_t^{(i)}$ before reaching a fully connected layer. To emulate the operation performed by STP, we consider a weight vector $\mathbf{w}$ for weighting the output of different kernels in the filter bank

$$y_k = f\left(\mathbf{w}^T \mathbf{h}_k\right) \tag{8}$$

where $\mathbf{h}_k$ concatenates the set of $h_k^{(i)}$ from different kernels i but the same input index k. By optimizing the kernels, similar CNNs can be trained to classify images [66, 67] as well as sounds [68, 69].

In a synapse with STP, the discretized efficacy train of the *ith* afferent, $\mathbf{e}_t^{(i)}$, results from a convolution and a nonlinear readout of the discretized spike train $\mathbf{s}_t^{(i)}$

$$e_t^{(i)} = f\left(\mathbf{k}_i^T \mathbf{s}_{t:t+K}^{(i)}\right) \mathbf{s_t}^{(i)}, \tag{9}$$

which is simply the discretized version of Eq. 2 and 3. This discretization makes clear the parallel with a convolutional layer in Eq. 7. As the spike train is conceived as a stochastic random variable sampling a potential [34, 48, 49], the stochastic spike train is

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

analogous to the dropout mask $\mathbf{m}$. The efficacy train triggers PSCs, which are pooling the efficacy train on the PSC : $e_k^{(i)} = \sum_{t=k}^{k+Z} \epsilon_{t-k} e_t^{(i)}$, where ϵ_i is a discretized and normalized PSC. Then, different synaptic afferents, with possibly different efficacy kernels (Fig. 8), are combined via by synaptic weights before taking a nonlinear readout at the cell body [34, 49] or the dendrites [70] to give rise to an instantaneous firing rate ρ_t:

$$\rho_t = f\left(\mathbf{w}^T \mathbf{e}_t\right) \tag{10}$$

This equation corresponds to the fully connected layer that followed a pooling operation, Eq. 8. Together, we find a striking parallel between the formalism developed here to describe STP and that of an artificial neural network which uses dropout at a convolutional layer followed by a fully connected layer.

Discussion

The linear-nonlinear framework has been able to capture core elements of subcellular [47], cellular [34, 36, 37, 71] and network signalling. We have shown that the same framework aptly captures synaptic dynamics. In the SRP framework, activity-dependent changes in efficacy are captured by an efficacy kernel. We have shown that switching the polarity of the kernel captures whether STD or STF is observed. Extending previous work at ribbon synapses [72], we have shown that the modeling framework captures multiple experimental features of synaptic dynamics. The model successfully reproduces the experimental extracellular calcium concentration manipulations seen to affect high-frequency stimulation responses. The framework can naturally capture long-lasting effects such as post-burst facilitation. Finally, by considering the dynamics of stochastic properties, a maximum likelihood approach can estimate model parameters from complex, time-limited, and physiological stimulation patterns. The added flexibility and the efficient inference are of interest to large scale characterization of synaptic dynamics [73] as well as the understanding information processing of neural networks [15, 74].

When summarizing dynamic properties with two time-dependent functions we called kernels, one is compelled to ask, what is their biophysical implementation? By analogy with characterization of neuronal excitability, the answer is likely to involve a mixture of independent mechanisms. The membrane kernel, for instance, depends on membrane resistance and membrane capacitance, but also the density of low-threshold channels, such as A- and H-type currents. Similarly, the efficacy kernel is likely to reflect residual presynaptic calcium concentration, the changing size of the readily releasable pool [31] but also many other possible mechanisms. Determining the relative importance of these processes, however, is not possible with the methodology described here. This could be achieved only with a combination of experiments aimed at isolating independent mechanisms and a detailed biophysical model, at the cost of constructing a model with reduced predictive power. In our view, the modeling framework presented here is less a tool for identifying molecular mechanisms, but rather one for the characterization, network simulations, and theoretical analysis [25, 75, 76] of the diversity of synaptic dynamics across signalling pathways [17], cell types [14, 50] or subcellular compartments [77].

There remains limitations to this approach; one such limitation is the choice of a gamma distribution of release sizes. Formally, this modeling choice means that the model replaces release failures with small to very small releases. In other terms, whereas the presence of release failures is a bimodal or multimodal distribution of amplitudes, the model assumes that the distribution of evoked amplitudes is unimodal. Nonetheless, recent work has shown that the release size distribution appears unimodal despite being

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

generated by multiple modes [56]. We have argued that for the small vesicle sizes at central synapses, quantal peaks are smeared by quantal variability [56]. When considering electrophysiological preparations where multiple synapses are simultaneously activated [27, 58, 78], the diversity of synaptic weights will strengthen further the assumption for a gamma-distribured, right-skewed and unimodal distribution.

Another unanswered question is that having explored various monotonic progressions of variability, can the model capture a non-monotonic progression? This case is relevant because the random and equally likely release of a number of vesicles will give rise to a non-monotonic progression of variability when release probability is changing over a larger range. For instance, in a facilitating synapse where multiple release sites increase an initially low release probability through a high-frequency train, the variability will first increase and then decrease. This convex, non-monotonic progression arises from the fact that variability is at its lowest point either when release probability is zero or when it is one. Given the mathematical features of the model, it may be possible to generate such a non-monotonic progression of variability with a biphasic σ-kernel.

Previous modeling and experimental work has established that dendritic integration can follow a hierarchy of linear-nonlinear processing steps [47, 70, 79]. Subcellular compartments filter and sum synaptic inputs through an integration kernel encapsulating a local passive and quasi-active properties. Active properties are responsible for a static nonlinear readout and for communication toward the cell body. Much in the same spirit, the work presented here extends this model by one layer, where presynaptic spikes first pass through a linear-nonlinear step before entering dendrites (Fig. 8). Since synapses at different locations or from different pathways may have different synaptic dynamics [17, 77], and since spiking neural codes can multiplex streams of information [8, 80, 81], these synaptic properties have the capability to extract different streams of information from multiple pathways and to process these possibly independent signals in segregated compartments.

The structure of information processing arising from this picture bears a striking resemblance with multi-layer convolutional neural networks [82, 83]. But it should be noted that the convolution takes place along the temporal dimension instead of the spatial dimension for many neural network applications. Yet, this algorithmic similarity suggests that the linear-nonlinear structure of synaptic processing capabilities on neural and neuronal networks. Whether the STP is controlled by genes [84], activity-dependent plasticity [85, 86], retrograde signalling [87], or neuromodulation [88, 89], a particular choice of efficacy kernels, when combined with a nonlinear readout, can optimize information processing.

Methods

All numerical simulations and parameter inference were done in Python using the numpy and scipy packages [90, 91].

Tsodyks-Markam Model and its Modifications

The Tsodyks-Markram (TM) model was first presented in 1997 [24] as a phenomenological model of depressing synapses between cortical pyramidal neurons and was quickly extended to account for short-term facilitating synapses [11, 50]. In the TM model, the normalized PSC amplitude μ_n at a synapse caused by spike n of a presynaptic spike train is defined as

$$\mu_n = R_n u_n \tag{11}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

where two factors u_n and R_n describe the utilized and recovered efficacy of the synapse, respectively. The temporal evolution of these variables are described by the following ordinary differential equations:

$$\frac{dR(t)}{dt} = \frac{1 - R(t)}{\tau_R} - u(t^-)R(t^-)S(t) \tag{12}$$

$$\frac{du(t)}{dt} = \frac{U - u(t)}{\tau_u} + f\left[1 - u(t^-)\right]S(t) \tag{13}$$

where f is the facilitation constant, τ_u the facilitation time scale, U the baseline efficacy and τ_R the depression timescale. The spike-dependent changes in R and u are implemented by the dirac delta function within the spike train $S(t)$. The notation t^- indicates that the function should be evaluated as the limit approaching the spike times from below.

In the TM model, facilitation is modelled as spike-dependent increases in the utilized efficacy u. Immediately after each spike, the efficacy increases by $f(1 - u(t^-))$. This efficacy jump depends on a facilitation constant f and on the efficacy immediately before the spike $u(t^-)$. Therefore, as u increases during a spike train, the spike-dependent 'jump' decreases for each subsequent spike. As a consequence, TM models of facilitating synapses are limited to a logarithmically saturating – that is, sublinear – facilitation.

To allow supralinear facilitation, we introduce a small change in the spike-dependent increase of factor u:

$$\frac{du(t)}{dt} = \frac{U - u(t)}{\tau_u} + u(t^-)f\left[1 - u(t^-)\right]\delta\left(t - t_S\right). \tag{14}$$

In this new model, given a presynaptic spike train at constant frequency, the size of the spike-dependent jump $u(t^-)f\left[1 - u(t^-)\right]$ saturates logarithmically for $u > 0.5$ but is increasing exponentially while $u < 0.5$. Thus this model provides supralinear facilitation in the low efficacy regimen, and it switches to sublinear facilitation for larger efficacies.

These models can be integrated between two spikes n and $n + 1$, separated by time Δt to speed up the numerical implementation [50]. For the classic TM model we have

$$R_{n+1} = 1 - \left[1 - R_n\left(1 - u_n\right)\right]\exp\left(-\frac{\Delta t}{\tau_R}\right) \tag{15}$$

$$u_{n+1} = U + \left[u_n + f\left(1 - u_n\right) - U\right]\exp\left(-\frac{\Delta t}{\tau_u}\right) \tag{16}$$

Similarly, the generalized model introduced in this work can be integrated between spikes:

$$u_{n+1} = U + \left[u_n + f\left(1 - u_n\right)u_n - U\right]\exp\left(-\frac{\Delta t}{\tau_u}\right) \tag{17}$$

Where $u_n^+ = u_n + f\left(1 - u_n\right)u_n$ is the value of u after the spike-dependent increase following the nth spike. In both models, at time $t = 0$, we assume no previous activation, therefore $R_0 = 1$ and $u_0 = U$.

Statistical Inference

To extract the properties of the model from experimental data, we developed a maximum likelihood approach. Given a set of amplitudes $\mathbf{y} = \{y_1, y_2, ..., y_i, ..., y_n\}$ resulting from a stimulation spike-train S, we want to find the parameters θ maximizing the likelihood $p(\mathbf{y}|S, \theta)$. For the mathematical model presented here, the negative log-likelihood (NLL) is:

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$NLL(\mathbf{y}|S,\theta) = \sum_i \left[\frac{y_i\mu_i}{\sigma_i^2} - \left(\frac{\mu_i^2}{\sigma_i^2} - 1\right) \ln\left(\frac{y_i\mu_i}{\sigma_i^2}\right) + \ln\left(\frac{\Gamma(\frac{\mu_i^2}{\sigma_i^2})\sigma_i^2}{\mu_i}\right)\right] \tag{18}$$

where μ_i and σ_i are shorthand for efficacy and standard deviation at the ith spike time: $\mu_i = \mu(t_i)$, $\sigma_i = \sigma(t_i)$, that is, the elements of the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.

We parametrize the time-dependent standard deviation and mean of the gamma distribution by expanding the filters $\mathbf{k}_\mu$ and $\mathbf{k}_\sigma$ in a linear combination of nonlinear basis: $k_\mu(t) = \sum_l a_l h_l(t)$, and $k_\sigma(t) = \sum_m c_m h_m(t)$. Typical choices for such nonlinear basis are raised cosine [32], splines [63, 64], rectangular [92] or exponential decays [34]. In counterpart to the numerical simulations where the kernels are made of a combination of exponential functions with different decay time constants, we have used this choice of basis functions.

In this framework, hyper-parameters are the choice of the number of basis functions, $l \in [0, L]$ and $m \in [0, M]$, as well as the decay time scale for each basis function $h_l(t) = \Theta(t)e^{-t/\tau_l}/\tau_l$, where $\Theta(t)$ is a Heaviside function. Free parameters are the amplitude of the basis functions $\{a_l\}$, $\{c_m\}$ and the scaling factor σ_0. By chosing hyper-parameters *a priori*, the modeller must choose a number of bases that is neither too big to cause overfitting, nor too small to cause model rigidity. The choice of time constant is made to tile exhaustively the range of physiologically relevant time scales. It is important to note that, because a combination of exponential basis functions can be used to capture a decay time scale absent from the set of τ hyper-parameters, the choice of τ does not specify the time scale of synaptic dynamics. The time-scale will be determined by inferring the relative amplitude of the basis funcitons. We can label the baseline parameter as the coefficient regulating the amplitude of a constant basis function, such that $a_0 = bh_0(t) = b_\mu$ and $c_0 = b_\sigma h_0(t) = b_\sigma$. There are thus $L + 2 + M + 2$ free parameters in total :

$$\theta = \{a_0, ..., a_L, \sigma_0, c_0, ..., c_M\}$$

To perform parameter inference, we first filter the data using the set of basis functions and stored the filtered spike train just before each spike in a matrix. Each row of the matrix corresponds to an individual basis function, and each column corresponds to spike timings. The matrix, X, thus stores the result of the convolution between the various basis function (rows) and at the time of the various spikes (columns).

For simplicity, it is convenient to take the same choice of basis functions for the efficacy and the variance kernel. The amplitudes are expressed in a vector $\theta_\mu = \{a_0, ..., a_L\}$, for the efficacy kernel, and $\theta_\sigma = \{c_0, ..., c_M\}$ for the variance kernel. Using this matrix notation, the linear combination is expressed as a matrix multiplication:

$$\boldsymbol{\mu} = \frac{1}{f(a_0)} f\left(X^T \boldsymbol{\theta}_\mu\right)$$

$$\boldsymbol{\sigma} = \sigma_0 f\left(X^T \boldsymbol{\theta}_\sigma\right)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ have length n and can be used to evaluate the NLL according to Eq. 18. Performing a grid search of the parameter space around initialized parameter values, we can obtain the landscape for the function, and ascertain the presence of convexity. The inferred parameters will then be the set of θ_μ and θ_σ minimizing the NLL over the training set.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Author contributions

RN conceived the study. RN, DT and JR performed the analysis. KT contributed the experimental data. RN and KT supervised the project. All co-authors contributed to writing the article.

Funding

NSERC Discovery Grant 06972 (RN). CIHR Project Grant RN38364 (RN and KT). Neurasmus EMJMD scholarship (JR).

Acknowledgments

We thank Alexandre Payeur, Ezekiel Williams, Anup Pilail, Emerson Harkin and Jean-Claude Béïque for helpful comments.

References

1. Feng T. Studies on the neuromuscular junction. XXVI. The changes of the end-plate potential during and after prolonged stimulation. Chinese Journal of Physiology. 1941;16:341–372.
2. Eccles JC, Katz B, Kuffler SW. Nature of the" endplate potential" in curarized muscle. Journal of Neurophysiology. 1941;4(5):362–387.
3. Magleby K, Zengel JE. A dual effect of repetitive stimulation on post-tetanic potentiation of transmitter release at the frog neuromuscular junction. The Journal of Physiology. 1975;245(1):163–182.
4. Varela JA, Sen K, Gibson J, Fost J, Abbott L, Nelson SB. A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. J Neurosci. 1997;17(20):7926–7940.
5. Zucker RS, Regehr WG. Short-term synaptic plasticity. Annual review of physiology. 2002;64(1):355–405.
6. Neubrandt M, Oláh VJ, Brunner J, Marosi EL, Soltesz I, Szabadics J. Single bursts of individual granule cells functionally rearrange feedforward inhibition. Journal of Neuroscience. 2018;38(7):1711–1724.
7. Crick F. Function of the thalamic reticular complex: the searchlight hypothesis. Proc Natl Acad Sci USA. 1984;81(14):4586–4590.
8. Naud R, Sprekeler H. Sparse bursts optimize information transmission in a multiplexed neural code. Proceedings of the National Academy of Sciences. 2018;doi:10.1073/pnas.1720995115.
9. Reyes A, Lujan R, Rozov A, Burnashev N, Somogyi P, Sakmann B. Target-cell-specific facilitation and depression in neocortical circuits. Nat Neurosci. 1998;1(4):279–285.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

10. Scanziani M, Gähwiler BH, Charpak S. Target cell-specific modulation of transmitter release at terminals from a single axon. Proceedings of the National Academy of Sciences. 1998;95(20):12004–12009.

11. Markram H, Wu Y, Tosdyks M. Differential signaling via the same axon of neocortical pyramidal neurons. Proc Natl Acad Sci USA. 1998;95:5323–5328.

12. De Pasquale R, Sherman SM. Synaptic properties of corticocortical connections between the primary and secondary visual cortical areas in the mouse. J Neurosci. 2011;31(46):16494–16506.

13. Sherman SM. Thalamocortical interactions. Current opinion in neurobiology. 2012;22(4):575–579.

14. Pala A, Petersen CC. In vivo measurement of cell-type-specific synaptic connectivity and synaptic transmission in layer 2/3 mouse barrel cortex. Neuron. 2015;85(1):68–75.

15. Ghanbari A, Malyshev A, Volgushev M, Stevenson IH. Estimating short-term synaptic plasticity from pre-and postsynaptic spiking. PLoS computational biology. 2017;13(9):e1005738.

16. Ghanbari A, Ren N, Keine C, Stoelzel C, Englitz B, Swadlow HA, et al. Modeling the short-term dynamics of in vivo excitatory spike transmission. Journal of Neuroscience. 2020;40(21):4185–4202.

17. Granseth B, Ahlstrand E, Lindström S. Paired pulse facilitation of corticogeniculate EPSCs in the dorsal lateral geniculate nucleus of the rat investigated in vitro. The Journal of physiology. 2002;544(2):477–486.

18. Felmy F, Neher E, Schneggenburger R. Probing the intracellular calcium sensitivity of transmitter release during synaptic facilitation. Neuron. 2003;37(5):801–811.

19. Lefort S, Petersen CC. Layer-Dependent Short-Term Synaptic Plasticity Between Excitatory Neurons in the C2 Barrel Column of Mouse Primary Somatosensory Cortex. Cerebral Cortex. 2017;27(7):3869–3878.

20. Wang Y, Markram H, Goodman PH, Berger TK, Ma J, Goldman-Rakic PS. Heterogeneity in the pyramidal network of the medial prefrontal cortex. Nature neuroscience. 2006;9(4):534.

21. Savanthrapadian S, Meyer T, Elgueta C, Booker SA, Vida I, Bartos M. Synaptic properties of SOM-and CCK-expressing cells in dentate gyrus interneuron networks. Journal of Neuroscience. 2014;34(24):8197–8209.

22. Chamberland S, Evstratova A, Tóth K. Interplay between synchronization of multivesicular release and recruitment of additional release sites support short-term facilitation at hippocampal mossy fiber to CA3 pyramidal cells synapses. Journal of Neuroscience. 2014;34(33):11032–11047.

23. Hennig MH. Theoretical models of synaptic short term plasticity. Frontiers in computational neuroscience. 2013;7:45.

24. Tsodyks M, Markram H. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. Proc Natl Academy of Sci, USA. 1997;94:719–723.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

25. Tsodyks M, Pawelzik K, Markram H. Neural Networks with Dynamic Synapses. Neural Computation. 1998;10(4):821–835.

26. Costa RP, Sjöström PJ, Van Rossum MC. Probabilistic inference of short-term synaptic plasticity in neocortical microcircuits. Front Comput Neurosci. 2013;7.

27. Barri A, Wang Y, Hansel D, Mongillo G. Quantifying repetitive transmission at chemical synapses: a generative-model approach. eNeuro. 2016;3(2):ENEURO–0113.

28. Barroso-Flores J, Herrera-Valdez MA, Galarraga E, Bargas J. Models of Short-Term Synaptic Plasticity. In: The Plastic Brain. Springer; 2017. p. 41–57.

29. Fuhrmann G, Cowan A, Segev I, Tsodyks M, Stricker C. Multiple mechanisms govern the dynamics of depression at neocortical synapses of young rats. The Journal of physiology. 2004;557(2):415–438.

30. Chamberland S, Timofeeva Y, Evstratova A, Volynski K, Tóth K. Action potential counting at giant mossy fiber terminals gates information transfer in the hippocampus. Proceedings of the National Academy of Sciences. 2018;115(28):7434–7439.

31. Kobbersmed JR, Grasskamp AT, Jusyte M, Böhme MA, Ditlevsen S, Sørensen JB, et al. Rapid regulation of vesicle priming explains synaptic facilitation despite heterogeneous vesicle: Ca2+ channel distances. eLife. 2020;9.

32. Pillow J, Paninski L, Uzzell V, Simoncelli E, Chichilnisky E. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. Journal of Neuroscience. 2005;25(47):11003–11013.

33. Pillow J, Shlens J, Paninski L, Sher A, Litke A, Chichilnisky E, et al. Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature. 2008;454(7207):995–999.

34. Mensi S, Naud R, Avermann M, Petersen CCH, Gerstner W. Parameter Extraction and Classification of Three Neuron Types Reveals two Different Adaptation Mechanisms. Journal of Neurophysiology. 2012;107:1756–1775.

35. Gerstner W, Kistler W, Naud R, Paninski L. Neuronal Dynamics. Cambridge: Cambridge University Press; 2014.

36. Pozzorini C, Mensi S, Hagens O, Naud R, Koch C, Gerstner W. Automated high-throughput characterization of single neurons by means of simplified spiking models. PLoS Comp Biol. 2015;11(6):e1004275.

37. Teeter C, Iyer R, Menon V, Gouwens N, Feng D, Berg J, et al. Generalized leaky integrate-and-fire models classify multiple neuron types. Nature communications. 2018;9(1):709.

38. Maass W, Zador AM. Dynamic stochastic synapses as computational units. In: Advances in neural information processing systems; 1998. p. 194–200.

39. Oswald AMM, Lewis JE, Maler L. Dynamically interacting processes underlie synaptic plasticity in a feedback pathway. Journal of neurophysiology. 2002;87(5):2450–2463.

40. Chichilnisky EJ. A simple white noise analysis of neuronal light responses. Network. 2001;12(2):199–213.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

41. Paninski L. Maximum likelihood estimation of cascade point-process neural encoding models. Network: Computation in Neural Systems. 2004;15:243–262.

42. Truccolo W, Eden U, Fellows M, Donoghue JP, Brown EN. A point process framework for relating neural spiking activity to spiking history, neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. Journal of Neurophysiology. 2005;93:1074–1089.

43. Wu MCK, David SV, Gallant JL. Complete functional characterization of sensory neurons by system identification. Annu Rev Neurosci. 2006;29:477–505.

44. Ostojic S, Brunel N. From spiking neuron models to linear-nonlinear models. PLoS Comput Biol. 2011;7(1):e1001056. doi:10.1371/journal.pcbi.1001056.

45. McFarland JM, Cui Y, Butts DA. Inferring nonlinear neuronal computation based on physiologically plausible inputs. PLoS computational biology. 2013;9(7):e1003143.

46. Vintch B, Movshon JA, Simoncelli EP. A convolutional subunit model for neuronal responses in macaque V1. Journal of Neuroscience. 2015;35(44):14829–14841.

47. Ujfalussy BB, Makara JK, Lengyel M, Branco T. Global and Multiplexed Dendritic Computations under In Vivo-like Conditions. Neuron. 2018;100(3):579–592.

48. Gerstner W, van Hemmen JL. Associative memory in a network of 'spiking' neurons. Network. 1992;3:139–164.

49. Jolivet R, Rauch A, Lüscher H, Gerstner W. Predicting spike timing of neocortical pyramidal neurons by simple threshold models. Journal of Computational Neuroscience. 2006;21:35–49.

50. Markram H, Wang Y, Tsodyks M. Differential signaling via the same axon of neocortical pyramidal neurons. Proc Natl Acad Sci USA. 1998;95(9):5323–5328.

51. Vyleta NP, Jonas P. Loose coupling between Ca2+ channels and release sensors at a plastic hippocampal synapse. Science. 2014;343(6171):665–670.

52. Scott R, Rusakov DA. Main determinants of presynaptic Ca2+ dynamics at individual mossy fiber–CA3 pyramidal cell synapses. Journal of Neuroscience. 2006;26(26):7071–7081.

53. Larkman A, Hannay T, Stratford K, Jack J. Presynaptic release probability influences the locus of long-term potentiation. Nature. 1992;360(6399):70.

54. Fuhrmann G, Segev I, Markram H, Tsodyks M. Coding of temporal information by activity-dependent synapses. Journal of neurophysiology. 2002;87(1):140–148.

55. Lavoie N, Jeyaraju DV, Peralta MR, Seress L, Pellegrini L, Tóth K. Vesicular zinc regulates the Ca2+ sensitivity of a subpopulation of presynaptic vesicles at hippocampal mossy fiber terminals. Journal of Neuroscience. 2011;31(50):18251–18265.

56. Soares C, Trotter D, Longtin A, Béïque JC, Naud R. Parsing out the variability of transmission at central synapses using optical quantal analysis. bioRxiv. 2019; p. 624692.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

57. Bekkers J, Richerson G, Stevens C. Origin of variability in quantal size in cultured hippocampal neurons and hippocampal slices. Proceedings of the National Academy of Sciences. 1990;87(14):5359–5362.

58. Bhumbra GS, Beato M. Reliable evaluation of the quantal determinants of synaptic efficacy using Bayesian analysis. Journal of neurophysiology. 2013;109(2):603–620.

59. Hefft S, Kraushaar U, RP Geiger J, Jonas P. Presynaptic short-term depression is maintained during regulation of transmitter release at a GABAergic synapse in rat hippocampus. The Journal of physiology. 2002;539(1):201–208.

60. Loebel A, Le Bé JV, Richardson MJ, Markram H, Herz AV. Matched pre-and post-synaptic changes underlie synaptic plasticity over long time scales. Journal of Neuroscience. 2013;33(15):6257–6266.

61. Gerstner W, Naud R. How Good Are Neuron Models? Science. 2009;326:379–380.

62. Dobrunz LE, Stevens CF. Response of hippocampal synapses to natural stimulation patterns. Neuron. 1999;22(1):157–166.

63. Kass RE, Ventura V. A spike-train probability model. Neural Computation. 2001;13:1713–1720.

64. Gerhard F, Pipa G, Lima B, Neuenschwander S, Gerstner W. Extraction of network topology from multi-electrode recordings: is there a small-world effect? Frontiers in computational neuroscience. 2011;5:4.

65. McCullagh P, Nelder JA. Generalized linear models. vol. 37. 2nd ed. Chapman & Hall/CRC; 1998.

66. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research. 2014;15(1):1929–1958.

67. LeCun Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision. In: Proceedings of 2010 IEEE international symposium on circuits and systems. IEEE; 2010. p. 253–256.

68. Dahl GE, Sainath TN, Hinton GE. Improving deep neural networks for LVCSR using rectified linear units and dropout. In: 2013 IEEE international conference on acoustics, speech and signal processing. IEEE; 2013. p. 8609–8613.

69. Zeghidour N, Xu Q, Liptchinsky V, Usunier N, Synnaeve G, Collobert R. Fully convolutional speech recognition. arXiv preprint arXiv:181206864. 2018;.

70. Payeur A, Béïque JC, Naud R. Classes of dendritic information processing. Current opinion in neurobiology. 2019;58:78–85.

71. Pillow JW, Paninski L, Simoncelli EP. Maximum Likelihood estimation of a stochastic integrate-and-fire model. In: Thrun S, Saul L, Schölkopf B, editors. Advances in Neural Information Processing Systems. vol. 16; 2004. p. 1311–1318.

72. Schröder C, James B, Lagnado L, Berens P. Approximate bayesian inference for a mechanistic model of vesicle release at a ribbon synapse. In: Advances in Neural Information Processing Systems; 2019. p. 7068–7078.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

73. Lee JH, Campagnola L, Seeman SC, Jarsky TH, Mihalas SH. Functional synapse types via characterization of short-term synaptic plasticity. bioRxiv. 2019; p. 648725.

74. Ghanbari A, Ren N, Keine C, Stoelzel C, Englitz B, Swadlow H, et al. Functional connectivity with short-term dynamics explains diverse patterns of excitatory spike transmission in vivo. bioRxiv. 2018; p. 475178.

75. Aitchison L, Pouget A, Latham PE. Probabilistic synapses. arXiv preprint arXiv:14101029. 2014;.

76. Schmutz V, Gerstner W, Schwalger T. Mesoscopic population equations for spiking neural networks with synaptic short-term plasticity. The Journal of Mathematical Neuroscience. 2020;10(1):1–32.

77. Grillo FW, Neves G, Walker A, Vizcay-Barrena G, Fleck RA, Branco T, et al. A distance-dependent distribution of presynaptic boutons tunes frequency-dependent dendritic integration. Neuron. 2018;99(2):275–282.

78. Barros Zulaica N, Rahmon J, Chindemi G, Perin R, Markram H, Ramaswamy S, et al. Estimating the Readily-Releasable Vesicle Pool Size at Synaptic Connections in a Neocortical Microcircuit. Frontiers in synaptic neuroscience. 2019;11:29.

79. Larkum M, Nevian T, Sandler M, Polsky A, Schiller J. Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. Science. 2009;.

80. Kayser C, Montemurro M, Logothetis N, Panzeri S. Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. Neuron. 2009;61(4):597–608.

81. Herzfeld DJ, Kojima Y, Soetedjo R, Shadmehr R. Encoding of action by the Purkinje cells of the cerebellum. Nature. 2015;526(7573):439.

82. Lecun Y, Bengio Y. Convolutional networks for images, speech, and time-series. In: The handbook of brain theory and neural networks. MIT Press; 1995.

83. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems; 2012. p. 1097–1105.

84. Sylwestrak EL, Ghosh A. Elfn1 regulates target-specific release probability at CA1-interneuron synapses. Science. 2012;338(6106):536–540.

85. Senn W, Tsodyks M, Markram H. An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timing. Neural Computation. 2001;13:35–67.

86. Costa RP, Padamsey Z, D'Amour JA, Emptage NJ, Froemke RC, Vogels TP. Synaptic transmission optimization predicts expression loci of long-term plasticity. Neuron. 2017;96(1):177–189.

87. Sjöström PJ, Turrigiano GG, Nelson SB. Multiple forms of long-term plasticity at unitary neocortical layer 5 synapses. Neuropharmacology. 2007;52(1):176–184.

88. Ding S, Li L, Zhou FM. Presynaptic serotonergic gating of the subthalamonigral glutamatergic projection. Journal of Neuroscience. 2013;33(11):4875–4885.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.133892; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

89. Takkala P, Woodin MA. Muscarinic acetylcholine receptor activation prevents disinhibition-mediated LTP in the hippocampus. Frontiers in cellular neuroscience. 2013;7:16.

90. Oliphant T. NumPy: A guide to NumPy; 2006–. USA: Trelgol Publishing. Available from: `http://www.numpy.org/`.

91. Jones E, Oliphant T, Peterson P, et al.. SciPy: Open source scientific tools for Python; 2001–. Available from: `http://www.scipy.org/`.

92. Pozzorini C, Naud R, Mensi S, Gerstner W. Temporal whitening by power-law adaptation in neocortical neurons. Nat Neurosci. 2013;16(7):942–948.