2-12-2019

Estimating Short-term Synaptic Plasticity from Pre- and Postsynaptic Spiking

Abed Ghanbari
University of Connecticut - Storrs, abedxghanbari@gmail.com

Estimating Short-term Synaptic Plasticity from Pre- and Postsynaptic Spiking

Abed Ghanbari, PhD

University of Connecticut 2019

Information processing in the nervous system critically depends on dynamic changes in the strength of these connections. Characterizing functional connectivity between spiking neurons is an essential first step toward understanding how the brain processes information at the single-cell and population levels. As a dynamical system, our brain changes its function in time and implements these dynamics through synaptic plasticity. Synaptic plasticity refers to the ability of each neuron in the neural network to change its influence on other neurons in time. Short-term synaptic plasticity (STP) refers to changes in synaptic strengths on timescales ranging from a few milliseconds to a few seconds. STP has been extensively studied in vitro by stimulating a presynaptic input with pulses of different frequencies and observing depression or facilitation in the postsynaptic potentials (PSPs) or currents (PSCs). STP is believed to underlie temporal filtering of inputs, gain control, network stability, sound localization by coincidence detection, and working memory. However, since recording PSPs/PSCs in vivo is challenging, STP has not been fully characterized in awake, behaving animals. Rather than observing PSP/PSCs directly, here we deduce STP parameters from spike observations alone. We model the short-term changes in the probability of a postsynaptic spike following a presynaptic spike. In cross-correlations between pre- and postsynaptic spiking, monosynaptic connections show a rapid, transient change in the probability of evoking a postsynaptic spike, at a short delay after the presynaptic spike (~ <4 ms). We developed a model-based approach for decomposing the short-term changes in the probability of a postsynaptic spike into four components: 1) short-term synaptic plasticity, 2) integration of PSPs, 3) history effects from previous postsynaptic spikes, and 4) slow common input to both pre- and postsynaptic neurons. The observed spike probability depends on each of these factors as well as the synaptic strength itself and the distribution of presynaptic spike times. We developed an extension of a typical generalized linear model (GLM) to use only pre- and postsynaptic spike observations. Our dynamical GLM allows us to characterize short-term synaptic dynamics of a wide range of synaptic behaviors in vivo.

Estimating Short-term Synaptic Plasticity from Pre- and Postsynaptic Spiking

Abed Ghanbari

B.S., Amirkabir University of Technology, 2006
M.S., Sharif University of Technology, 2011

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2019

2019

APPROVAL PAGE

Doctor of Philosophy Dissertation

Estimating Short-term Synaptic Plasticity from Pre- and Postsynaptic Spiking

Presented by

Abed Ghanbari

Major Advisor

_____

Ian Stevenson

Associate Advisor

_____

Maxim Volgushev

Associate Advisor

_____

Monty A. Escabi

Associate Advisor

_____

Heather Read

Associate Advisor

_____

Sabato Santaniello

University of Connecticut

2019

Dedication

The opportunity to learn from my advisor has been an extraordinary gift. I am indebted to Ian Stevenson for being the best advisor anyone could ever imagine. His incredible depth and breadth of knowledge in grasping novel ideas and his confidence in me were essential in my academic development. To Davide Zoccolan for his life-changing trust in me and bringing me to the neuroscience world. I am deeply grateful to my wonderful co-authors – Maxim Volgushev, Harvey Swadlow, Bernhard Englitz, Christian Keine, and Naixin Ren– for being passionate scientists. I learned a lot from them, and I am looking forward to collaborating with them on many more projects to come.

I am truly thankful to my parents, Monireh and Seifollah; my sisters, Mahdieh and Fatemeh; my nephew Abtin; my cousins, Hosein, Hasan, Mohammad, Rasul, and AmirHossein; and my grandparents for their unconditional love. They made my childhood beautiful and they never stopped loving me.

To Laurie and Phillip Best who became my compass when the waves crashed right on to my boat and I was losing my grip.

To all my friends whom I spent time during my journey at UConn, who together we watched movies, camped, played volleyball and football, talked, walked, hiked, debated, smiled, and shared moments, feelings and stories. You all are the sole reminder of life and spending time with you brings joy and happiness in me. To name a few (in chronological order of our friendship): Armin, Leila, Hamed, Fatemeh, SinaK, PouyaA, AmirAli, Vanessa, Swing, Saman, Sajad, RezaN, MZ, Mahdis, Mona, Mani, Sevan, Sevana, Pouya, Mina, AT, AmirE, RezaA, Mohammad, HamidReza, Bahareh, MajidM, Bahar, Nasim, Eric, SinaT, Behnoush, Sara, Fariba, Hakimeh, Mehdi, Veronica, Nora, Marcella, AmirA, Rahim, Sahar, Lauren, and Helia– Thank you all for making me much happier.

Table of Contents

Chapter 1: Introduction

Neurons in our brain are connected through synaptic structures with varying strength in connectivity which allows one neuron to communicate selectively with another. Information processing in the nervous system critically depends on dynamic changes in the strength of these connections between neurons. Characterizing functional connectivity between spiking neurons is an essential first step toward understanding how the brain processes information at the single-cell and population levels to carry out neural computations. As a dynamical system, our brain changes its function in time and implements these dynamics through synaptic plasticity. Synaptic plasticity refers to the ability of each neuron in the neural network to change its influence on other neurons in time. Plasticity could refer to long term plasticity (LTP) which these changes take place in the scales of minutes and hours and is believed to be involved in learning and memory. On the other hand, short-term synaptic plasticity (STP) is the activity-dependent change of synaptic strength based on the history of presynaptic spiking activity [1]. STP allows the connections between neurons to evolve on timescales of tens to hundreds of milliseconds. Both of these plasticity forms have been observed in different regions of the brain and can directly affect synaptic transmission and act to alter information processing in neuronal circuits [2–4].

Short-term plasticity occurs due to calcium and vesicle dynamics in the synapse and varies across both cell-types and brain regions [5,6]. Based on their effects on synaptic efficacy, there are two types of STP, facilitation, where synaptic efficacy increases with consecutive presynaptic spikes, and depression, where synaptic efficacy decreases. Previous studies have shown that the STP

allows differences in temporal filtering [7], network stability [7], working memory [8] and regulation of plasticity on longer timescales [9]. These studies suggest that neural computation and higher-order cognitive functions are related to the neural network activity in short time-scales. Therefore, studying information transmission in short time-scales could lead to an understanding of the underlying neural computations.

Traditionally, synaptic transmission and short-term plasticity is studied using intracellular recordings where the effect of presynaptic spikes is directly observed in the postsynaptic potentials or currents. With intracellular recordings, electrophysiologists have precise control of the stimulus and are able to study how pharmacological manipulations alter plasticity. In many experiments, short-term changes in the amplitude of the PSPs/PSCs is summarized using the paired-pulse ratio (PPR) or the steady-state response to a train of stimuli. However, most of these controlled current injections assume that the synapse is in its resting state and does not account for the variability of postsynaptic response to arbitrary presynaptic spiking.

Using these intracellular recordings and to understand how different presynaptic spiking patterns result in changes in amplitude of postsynaptic potentials (PSPs), Tsodyks and Markram (TM) introduced a dynamical system of equations to explain the changes in release probability associated with short-term facilitation and the depletion of the resources in short-term depression [10,11]. The biophysical representation of the synaptic behavior in TM made it popular among others [12–14] and was used in several *in vitro* studies to fit intracellular recordings and to assess synaptic dynamics in different regions of the brain. However, large-scale intracellular recordings similar to in vitro experiment in behaving animals are currently challenging [15,16]. On the other hand, evidence of synaptic connectivity appears not only in postsynaptic potentials and currents but also in spike statistics and extracellular recordings might be used to study STP using spikes.

Large-scale extracellular recording techniques are growing [17] where hundreds of simultaneously recorded neurons and their synapses could be potentially studied simultaneously using their spiking activity. Previously, using the cross-correlation between pre- and postsynaptic spikes several studies have found evidence of synaptic connections [18,19]. They showed how it is possible to detect monosynaptic connections in cross-correlograms when postsynaptic spikes tend to follow presynaptic spikes with fast onsets and short latency [20]. To study STP using extracellular recordings, a descriptive method extended these methods by splitting cross-correlograms into short and long presynaptic inter-spike intervals (ISI). Using split cross-correlation approach, they inferred the STP type (i.e. facilitation and depression) based on the efficacy differences in short and long ISIs [21,22]. Although the split cross-correlation approach made it possible to observe STP *in vivo*, it only provides a qualitative description of synaptic dynamics rather than a detailed account of release probability and vesicle depletion dynamics for full sequence of presynaptic spikes.

Here we propose modelling approaches to study short-term synaptic plasticity using extracellular recordings of spiking activity. Our approach is able to include many covariates relating to network functions (i.e. common input, post-spike history, postsynaptic excitability) and could overcome the limitations of descriptive statistics (cross-correlations) in estimating biological meaningful parameters of a dynamical synapse.

Explicitly, we develop dynamical functional connectivity models that include short-term plasticity based on a generalized linear model (GLM). Our model predicts probability of postsynaptic spiking as a function of the observed pre- and postsynaptic spiking history [23]. Previous models of this type have revealed the underlying structure of retina [24] and cortex [25], and result in more accurate detection of weak synapses [26]. However, in these studies strength of the functional

connections is assumed to be static. Here we modify the GLM and its covariates to allow the effect of the presynaptic spikes on postsynaptic spiking probability to vary on short timescales. Our dynamical model connectivity will allow us to infer (i) how neural correlation changes in short timescale, (ii) how this change is related to short-term synaptic plasticity, and (iii) how accurately it can predict postsynaptic spiking probability.

In Chapter 2, we assess the ability of models to infer different forms of STP and validate these models using results from controlled *in vitro* experiments where we inject a presynaptic current as a combined current of a population of presynaptic neurons with simulated and known plasticity. Our results show that our models of dynamical functional connectivity can accurately recover the synaptic dynamics underlying spiking. Moreover, we examine the effects of synaptic weights, adaptation, stochastic vesicle release, spike sorting errors, and common input. These results suggest the feasibility of studying STP using spike observations.

In Chapter 3, we follow the *in vitro* experiment and use extracellular recordings from behaving animal to study STP under different tasks and brain regions. Similarly, here we deduce short-term dynamics of synaptic transmission from spike observations using an extension to generalized linear models (GLMs) by including multiple factors that regulate the spike transmission probability *in vivo*. The main challenge *in vivo* is the common input and presence of multiple inputs from other neurons. In three identified synapses, our model captures short-term dynamics of spike transmission and decomposes the effects of local patterns of pre- and postsynaptic spikes. We used this model (i) to estimate how spike transmission dynamically shifts depending on stimulus type, and (ii) to study how spike transmission is postsynaptic cell-type specific in a large-scale extracellular recording.

Together, these two chapters demonstrate how it may be possible to estimate short-term synaptic dynamics from spikes alone, without intracellular observations of membrane potentials or currents. Our exploration through *in vitro* experiments and *in vivo* data is an attempt to bring insight to study short-term dynamics of spike transmission and provides a potential framework to understand neural computation that underlies several aspects of network functions in short-term timescales in behaving animals.

Chapter 2 (Estimating short-term synaptic plasticity from pre- and postsynaptic spiking: a simulation and in vitro validation study) represents work already published as:

"Estimating short-term synaptic plasticity from pre- and postsynaptic spiking"

PLoS Computational Biology 13(9): e1005738.

In addition to the dissertation author, the following individuals also made contributions to the writing contained in this chapter:

Ian H. Stevenson[1,3]

Aleksey Malyshev[2]

Maxim Volgushev[3]

Affiliations:

1) University of Connecticut, Department of Biomedical Engineering

2) Institute of Higher Nervous Activity and Neurophysiology, Russian Academy of Science

3) University of Connecticut, Department of Psychological Sciences

Chapter 2: Estimating short-term synaptic plasticity from pre- and postsynaptic spiking: *a simulation and in vitro validation study*

Short-term synaptic plasticity (STP) critically affects the processing of information in neuronal circuits by reversibly changing the effective strength of connections between neurons on time scales from milliseconds to a few seconds. STP is traditionally studied using intracellular recordings of postsynaptic potentials or currents evoked by presynaptic spikes. However, STP also affects the statistics of postsynaptic spikes. Here we present two model-based approaches for estimating synaptic weights and short-term plasticity from pre- and postsynaptic spike observations alone. We extend a generalized linear model (GLM) that predicts postsynaptic spiking as a function of the observed pre- and postsynaptic spikes and allow the connection strength (coupling term in the GLM) to vary as a function of time based on the history of presynaptic spikes. Our first model assumes that STP follows a Tsodyks-Markram description of vesicle depletion and recovery. In a second model, we introduce a functional description of STP where we estimate how the coupling term is modified by directly inferring a function of presynaptic inter-spike intervals. To validate the models, we test the accuracy of STP estimation using the spiking of pre- and postsynaptic neurons with known synaptic dynamics. We first test our models using the responses of layer 2/3 pyramidal neurons to simulated presynaptic input with different types of STP, and then use simulated spike trains to examine the effects of spike-frequency adaptation, stochastic vesicle release, spike sorting errors, and common input. We find that, using only spike observations, both model-based methods can accurately reconstruct the time-varying synaptic weights of presynaptic inputs for different types of STP. Our models also capture

the differences in postsynaptic spike responses to presynaptic spikes following short vs long inter-spike intervals, similar to results reported for thalamocortical connections. These models may thus be useful tools for characterizing short-term plasticity from multi-electrode spike recordings *in vivo*.

Introduction

Short-term synaptic plasticity (STP) refers to fast and reversible changes of synaptic strength caused by the recent history of presynaptic spiking activity [1]. STP occurs on timescales from milliseconds to tens of seconds, and includes mechanisms for both facilitation of transmitter release, where synaptic strength increases with consecutive presynaptic spikes, and depression, where synaptic strength decreases. Facilitation and depression are mediated by the dynamics of presynaptic calcium and the depletion and replenishment of vesicles in the presynaptic terminals [1]. The relative contribution of facilitation and depression varies across synapses, cell types, and brain regions [5,6] with facilitation dominating at some synapses and depression at others. By shaping postsynaptic responses evoked by trains of presynaptic action potentials, STP alters neuronal information processing [2–4]. In vitro studies have shown that STP has profound effects on temporal filtering [7], network stability [7], and working memory [8]. Moreover, there is bidirectional interaction between STP and long-term synaptic changes: STP can determine the magnitude of long-term plasticity [27–30], and long-term synaptic changes also modify STP [30,31]. This results in an interplay between STP and long-term plasticity on multiple timescales [9,31]. Therefore, characterization of short-term plasticity in different systems is crucial for understanding neural computations.

Traditionally, short-term plasticity is studied using intracellular recordings where responses of the postsynaptic neuron to presynaptic stimulation are directly measured as evoked postsynaptic potentials or currents. Based on results of intracellular recordings Tsodyks, Markram, and colleagues developed a computational model that describes STP in terms of dynamics of resources and their utilization [11,32]. The Tsodyks-Markram (TM) model provides a phenomenological description of the short-term dynamics of synaptic responses in terms of 1) changes in the

probability of transmitter release (utilization), related to the dynamics of presynaptic calcium and, 2) the use and replenishment of synaptic vesicles (resources). The TM model accurately captures the dynamics of synaptic responses caused by STP, links the observed diversity in synaptic dynamics to differences in the model parameters (utilization, recovery of resources, and their time constants), and allows prediction of postsynaptic responses to an arbitrary sequence of presynaptic stimuli [33]. Although several alternative models of STP have been proposed [14,34], the TM model is the most broadly used because it provides a compact description of STP with biophysically relevant parameters.

The TM model had been successfully used in a number of intracellular studies to assess synaptic dynamics in different connections [33,35,36] and changes of synaptic dynamics induced by long-term plasticity [31], adaptation [37] or injury [38]. Traditionally TM model parameters are estimated from responses to presynaptic stimuli applied in bursts of different frequencies [11,31,37,38]. A recent study presented a Bayesian approach that estimates TM model parameters by fitting postsynaptic responses induced by stochastic trains of presynaptic spikes [33]. Thus, STP parameters can be extracted from responses to *in vivo*-like presynaptic activity. Here we ask whether it is possible to estimate STP parameters using only the spike trains of pre- and postsynaptic neurons without access to postsynaptic potentials or currents. If available such a method would greatly expand the possibilities for studying STP *in vivo*. Although multiple intracellular recordings or simultaneous extra and intracellular recordings *in vivo* are possible [18,39–41], they are technically prohibitive for large-scale studies. Techniques for large-scale extracellular recordings, on the other hand, allow simultaneous recording of spiking from hundreds of neurons [17,42,43]. Prior studies compared cross-correlograms calculated using presynaptic spikes occurring after short or long inter-spike intervals, and found evidence for both short-term

facilitation [44] and depression [20,21] of synaptic transmission *in vivo*. This split-correlograms approach, however, does not allow for a detailed reconstruction of synaptic weight for each presynaptic spike or estimation of underlying release probability and vesicular resources.

Here we develop two statistical methods that use pre- and postsynaptic spike trains to estimate the dynamics of short-term plasticity. Both approaches are based on a generalized linear model (GLM) that predicts postsynaptic spiking as a function of the observed pre- and postsynaptic spikes [23–25,45,46]. In these GLM-based methods we allow the effect of the presynaptic spikes to vary on short timescales as a function of the presynaptic spike timing. In a first model, the effect of presynaptic spikes is determined by the nonlinear dynamical equations of the TM model (TM-GLM). In a second model, we introduce a functional description of short-term plasticity based on a generalized bi-linear model (GBLM). Although the parameters in the second approach are no longer linked to biophysical properties the GBLM allows us to capture a wide range of neuronal interactions and synaptic dynamics.

To validate our models, we recorded spike responses of pyramidal neurons in vitro (cortical slices, layer 2/3 pyramids) to intracellularly injected currents composed of synaptic inputs with the known pre-defined short-term plasticity. We show that, using only pre- and postsynaptic spike trains, the TM-GLM can recover the underlying parameters of STP, and the GBLM is able to reconstruct synaptic dynamics using a descriptive plasticity "rule". Estimates provided by each of the two models were in good correspondence to ground truth values for a wide range of synaptic weights and time scales of facilitation and depression. Additionally, using simulated neurons we show that estimation of STP by these models is robust to several potential confounds: spike frequency adaptation, noise from probabilistic vesicle release, and spike sorting errors. The methods developed here, thus, have the potential to serve as powerful tools for large-scale studies of short-

term synaptic plasticity *in vivo*, including alterations of short-term plasticity during different behaviors, during learning, or as a result of pathology.

Methods and Models

Ethics Statement:

All animal use procedures conform to the principles outlined in the Guide for the Care and Use of Laboratory Animals (National Institutes of Health publication no. 86-23, revised 1985) and were approved by the Institutional Animal Care and Use Committee at the University of Connecticut.

A phenomenological generative spiking model of short-term plasticity:

Approaches using generalized linear models (GLMs) have proved to be effective tools for estimation neuronal connections from spike train data [47–49]. The standard GLM assumes that the spike train is a binary sequence of observations, $m(t)$, generated from a Poisson process. For a single pair of neurons, we model the conditional intensity, $\lambda(t)$, of this process as a linear combination of a baseline firing rate μ, a contribution from the presynaptic neuron $\boldsymbol{rx_t}$ and weighted contribution from the postsynaptic spike-history $\boldsymbol{sy_t}$ passed through an exponential nonlinearity (Fig. 2.1)

$$\begin{aligned} &\lambda(t \mid \mu, \boldsymbol{r}, \boldsymbol{s}) = \exp(\mu + \boldsymbol{rx_t} + \boldsymbol{sy}_t) \\ &m(t) \sim Poisson\big(\lambda(t \mid \mu, \boldsymbol{r}, \boldsymbol{s})\big) \\ &\boldsymbol{x_t} = [x_1(t), x_2(t), \dots, x_L(t)], \qquad x_j(t) = n(t) * b_j(t) \\ &\boldsymbol{y_t} = [y_1(t), y_2(t), \dots, y_L(t)], \qquad y_j(t) = m(t) * b_j(t) \end{aligned} \tag{1}$$

where $n(t)$ and $m(t)$ are the pre- and post- synaptic spike trains, respectively.

Our goal is to estimate the set of model parameters $\boldsymbol{r} = [\beta_c^{(1)}, \beta_c^{(2)}, \dots, \beta_c^{(L)}]$, $\boldsymbol{s} = [\beta_h^{(1)}, \beta_h^{(2)}, \dots, \beta_h^{(L)}]$ and μ, describing the coupling, $k(t)$, and post-spike filter, $h(t)$, which best predicts the postsynaptic firing $m(t)$.

$$k(t) = \sum_{j=1}^{L} \beta_c^{(j)} b_j(t);$$

$$h(t) = \sum_{j=1}^{L} \beta_h^{(j)} b_j(t); \quad (2)$$

$$b_j(t) = \frac{1}{2}\cos\left(\log\left(t + C_j\right) + \pi\right) + \frac{1}{2}$$

where $b_j(t)$ are raised-cosine basis functions which reduce dimensionality and allow a smooth representation of the two filters [24]. This stochastic model of a Poisson spiking neuron has a guaranteed convex log-likelihood which gives a unique set of parameters for its global maximum [46].

In order to model plasticity, we modified the GLM, allowing the contribution of coupling to vary over time. A conventional GLM treats all presynaptic spikes $n(t)$, equally, with each presynaptic spike having the same "weight" when influencing conditional intensity, $\lambda(t)$. To account for short-term facilitation and depression we modify the weights of each spike according to the phenomenological Tsodyks-Markram (TM) model [11]. The TM model describes the dynamics of resources R and their utilization U by the following system of differential equations:

$$\frac{dR(t)}{dt} = \frac{1 - R(t)}{D} - u(t^-)R(t^-)\delta(t - t_s)$$

$$\frac{du(t)}{dt} = \frac{U - u(t)}{F} + f[1 - u(t^-)]\delta(t - t_s) \quad (3)$$

where resources, $R(t)$, represent the portion of available vesicles which instantly decreases after each spike at t_s and gradually recovers with depression time constant D. The second equation describes release probability (utilization of resources), which instantly increases after each spike by $f[1 - u(t^-)]$, where f is the magnitude of facilitation and decays back to the baseline value, U, with facilitation time constant F.

The amplitude of the postsynaptic current $I_{syn}(t_i)$ evoked by presynaptic spike at t_i is described by

$$I_{syn}(t_i) = A\,R(t_i)u(t_i) \tag{4}$$

where A is the maximal current that can be evoked at that synapse if all resources are recovered ($R = 1$) and are released at once.

With different sets of parameters $\boldsymbol{\theta} = \{\boldsymbol{D}, \boldsymbol{F}, \boldsymbol{U}, \boldsymbol{f}\}$ this model can reproduce diverse types of short-term plasticity (depression, facilitation or a mixture of both) observed experimentally [33]. Using these dynamics, we create a "marked" point-process

$$n^*(t) = A\,R(t)u(t)n(t) \tag{5}$$

where $n^*(t)$ captures the amplitudes of PSCs at the time-points of presynaptic spikes and is zero otherwise. By using $n^*(t)$ instead of $n(t)$ in the modified GLM (TM-GLM), we account for STP in the coupling term. Note that when $n^*(t)$ is constant ($R(t)$ and $u(t)$ constant) the TM-GLM will describe a steady-state synapse with no short-term plasticity, and, in this case, TM-GLM is identical to the original GLM.

With the modified coupling term the original observation model is rewritten as

$$\begin{aligned} &\lambda(t \mid \mu, \boldsymbol{r}, \boldsymbol{s}) = \exp(\mu + \boldsymbol{r}\boldsymbol{x}_{\boldsymbol{t}}^* + \boldsymbol{s}\boldsymbol{y}_t) \\ &m(t) \sim Poisson\big(\lambda(t \mid \mu, \boldsymbol{r}, \boldsymbol{s})\big) \end{aligned} \tag{6}$$

Using the TM-GLM our goal is to estimate the static parameters of the synapse $\phi = \{\mu, \boldsymbol{r}, \boldsymbol{s}\}$, as well as the plasticity parameters $\boldsymbol{\theta} = \{\boldsymbol{D}, \boldsymbol{F}, \boldsymbol{U}, \boldsymbol{f}\}$, given the pre- and postsynaptic spike trains. Specifically, we aim to find maximum a posteriori (MAP) estimates of θ and ϕ that optimize

$$\begin{aligned} &p\left(\boldsymbol{\theta}, \boldsymbol{\phi} \middle| n_{post}(t), n_{pre}(t)\right) \\ &\qquad \propto p\left(n_{post}(t) \middle| \boldsymbol{\theta}, \boldsymbol{\phi}, n_{pre}(t)\right) p\left(\boldsymbol{\theta}, \boldsymbol{\phi} \middle| n_{pre}(t)\right) \end{aligned} \tag{7}$$

$$= p\left(n_{post}(t)\middle|\boldsymbol{\theta}, \boldsymbol{\phi}, n_{pre}(t)\right) p(\boldsymbol{\theta}) p(\boldsymbol{\phi})$$

To prevent over-fitting and assure nonnegative values, we introduce weakly informative priors on the plasticity parameters $p(\theta)$ to span the parameters space only over meaningful intervals and prevent the optimization from getting stuck at local minima. We then use coordinate ascent, maximizing the log posterior by alternating between optimizing the plasticity parameters given the GLM parameters and fitting the GLM given fixed plasticity parameters. Although this posterior is not guaranteed to be convex, in many cases, the non-convexity of GLM-like models does not lead to optimization problems [50–52]. Previous work estimating STP parameters from intracellular recordings suggests that, rather than point estimates, a fully Bayesian approach may provide a more accurate understanding of the parameters [33,53]. Although it is possible to use MCMC to sample from the posterior, the large number of function evaluations (compared to optimization) makes it less attractive for our model with spike observations.

When optimizing plasticity parameters (GLM parameters fixed) we randomly restart over the θ-space and use priors $\{0 < \boldsymbol{D}, \boldsymbol{F} < 2\} \sim gamma(\alpha = 1.2, \beta = 2)$ and $\{0 < \boldsymbol{U}, \boldsymbol{f} < 1\} \sim beta(1.01, 1.01)$. We optimize the plasticity parameters in the log-domain using two-metric projection and numerical differentiation of the posterior [54]. Additionally, we found that when optimizing the plasticity parameters, convergence is improved by normalizing the static coupling term $k(t)$ and optimizing an amplitude A (with prior $p(A) = cauchy(0, 50)$) alongside the parameters $\boldsymbol{\theta}$. These prior distributions and parameters were chosen to prevent the model from reaching the boundaries (e.g. U or f at 0 or 1), but they do introduce bias into the parameter estimates and may not necessarily work well for all sets of data.

When optimizing the static GLM parameters $\boldsymbol{\phi}$ (plasticity parameters fixed) we would typically assume $p(\boldsymbol{\phi})$ to be flat. However, we found that in some cases the coupling term $k(t)$ interacts

with the plasticity parameters. For instance, an excitatory depressing synapse will show a biphasic coupling term where a negative component can partially account for the reduced impact of a burst. To prevent this type of ambiguity we introduced a quadratic penalty on negative coupling coefficients $\boldsymbol{r}$ with the improper prior $\log p(r) \propto -\eta r^2 \mathbf{1}_{r<0}$. In practice, we use LBFGS optimization of the penalized log-likelihood and this ensures that the estimated coupling term is approximately positive for excitatory inputs and negative for inhibitory inputs. With limited data or when extending the model to multiple inputs additional types of regularization may be useful [48].

A nonparametric generalized bilinear model of STP:

The phenomenological model, described above, gives a clear view of the synaptic dynamics by searching over the $\boldsymbol{\theta}$-space of STP parameters. However, in cases were TM assumptions on synaptic dynamics such as vesicle release and changes of the calcium changes in presynaptic terminal doesn't hold, it may be preferable to have a model of STP that is not constrained to the TM dynamics. In a second type of model – the generalized bilinear model - instead of searching over the space of STP parameters we directly infer a short-term synaptic modification "rule". This generalized bilinear model (GBLM) compartmentalizes the coupling term into a stationary and a short-term plastic modification (Fig. 2.2).

$$\begin{aligned} \lambda(t \mid \mu, \boldsymbol{r}, \boldsymbol{s}, w(t)) &= exp(\mu + \boldsymbol{r}\boldsymbol{x}_t . w(t) + \boldsymbol{s}\boldsymbol{y}_t) \\ m(t) &\sim Poisson\big(\lambda(t \mid \mu, \boldsymbol{r}, \boldsymbol{s}, w(t))\big) \end{aligned} \tag{8}$$

Here the modification term, $w(t)$, weights the static coupling term depending on the history of presynaptic spiking. For a synapse with no plasticity, $w(t)$ equals to one and the coupling term, $\boldsymbol{r}\boldsymbol{x}_t$ is static and does not depend on previous presynaptic spiking. For a synapse with plasticity,

$w(t)$ >1 would increase the static coupling term $\boldsymbol{r}\boldsymbol{x}_t$ to account for facilitation and $w(t)$ <1 would decrease the $\boldsymbol{r}\boldsymbol{x}_t$ to account for depression. In both cases, the effect of w(t) on the coupling term decays with time and the coupling term recovers to its static form. We defined the modification function as:

$$w(t) \;=\; 1 + \sum_{k} q_k \sum_{l=0}^{T} \delta(t_k - l)\exp\left(-\frac{l}{\tau}\right) \qquad (9)$$

$$q_k = \boldsymbol{\alpha}\, \boldsymbol{B_m}(\Delta t_k)$$

where $q(\cdot)$ determines the amplitude of exponentially decaying effects from previous spikes on the synaptic weight. Here k indexes the presynaptic spikes with times t_k and previous inter-spike intervals Δt_k.

Although we could attempt to fit the decay function (instead of using single exponential) and its time-constant ($\tau = .2$) we fixed them to increase the robustness and speed of the maximum likelihood parameter fitting. Spikes are then convolved with the exponential kernel weighted by the modification terms $q(\cdot)$. To ensure $q(\cdot)$ is a smooth function we represent it using the B-spline bases, $B_m(\Delta t)$, with log-spaced sampling knots in Δt_k. The final model is linear in both the stationary parameters, $\{\mu, \boldsymbol{r}, \boldsymbol{s}\}$, and STP parameters, $\boldsymbol{q}$. To estimate the parameters, we alternate between two GLMs: fitting $\{\mu, \boldsymbol{r}, \boldsymbol{s}\}$ with fixed $\boldsymbol{q}$ and fitting $\boldsymbol{q}$ with fixed $\{\mu, \boldsymbol{r}, \boldsymbol{s}\}$ (both using iterative reweighted least squares - IRLS). Although the two GLMs are log-concave in this problem, the joint likelihood of $\{\mu, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{q}\}$ is not guaranteed to be concave. However, we find that in practice convergence is fast using the alternating method and random restarts results in the same final solution.

Experiments in Slices: Recording and Current Injection:

Slices of visual cortex were prepared from male Wistar rats (P21-P23) as described in detail in our prior work [55]. Extracellular solution used during preparation of slices and for perfusion of recording chamber contained (in mM): 125 NaCl, 2.5 KCl, 2 $CaCl_2$, 1 $MgCl_2$, 1.25 NaH_2PO_4, 25 $NaHCO_3$, 25 D-glucose and was bubbled with 95% O_2 and 5% CO_2. Patch clamp electrodes for whole cell recordings were filled with K-gluconate based solution (in mM: 130 K-Gluconate, 20 KCl, 4 Mg-ATP, 0.3 Na_2-GTP, 10 Na-Phosphocreatine, 10 HEPES) and had a resistance of 4–6 MΩ. Whole-cell recordings were made from layer 2/3 pyramidal neurons of rat visual cortex. Membrane potential responses to injection of fully-defined fluctuating current ([55]; see below) were recorded using the bridge mode of a Dagan BVC-700A amplifier (Dagan Corporation, USA). Data were digitized at 20 kHz (Digidata 1440A, Molecular Devices, USA) and stored for further processing. Timings of postsynaptic spikes were determined as positive-slope zero crossings of the membrane potential signal.

An artificial current for injection was designed to mimic the postsynaptic effect of a realistic cortical circuitry with inputs of different strength and unique short-term synaptic characteristics (Fig. 2.3). Current for injection was synthesized using a population of 96 presynaptic neurons (6 pools of 16 neurons, 8 excitatory and 8 inhibitory). Five sets of STP parameters were chosen to cover the whole spectrum of the plasticity from strong depression to strong facilitation. The sixth set of synapses did not express short-term plasticity. For each neuron, we generated an inhomogeneous Poisson spiking series with the log rate generated using a cubic spline function with 1 knot/s and standard normally distributed amplitudes. The rate is then scaled to generate an average spike rate of 5Hz and the spikes are weighted to generate the postsynaptic current amplitudes of the TM model. The weighted series of postsynaptic current amplitudes was then

convolved with a synaptic integration kernel to generate the artificial postsynaptic current traces. We generated the kernel as a difference of two exponentials with time constants of 1ms and 10ms. Eight different synaptic weights with a normal inverse cumulative distribution function ($\mu = .7$ & $\sigma = .93$) were used to create a pool of excitatory synapses. Same synaptic weights, but with a negative sign were used to generate currents produced by inhibitory neurons. Because the number and weight distributions for excitatory and inhibitory presynaptic neurons were the same, the total input current was balanced. We used 20 different realizations of the current for injection. The duration of each current trace was 46s. Injections of fluctuating currents were separated by intervals of 60-100s. The amplitude of the injected current was adjusted to produce membrane potential fluctuations of 10-15 mV. DC current was added to achieve the average postsynaptic firing rate of ~5Hz.

Thus, we knew the timing of presynaptic spikes for each simulated presynaptic neuron contributing a synaptic connection as well as its amplitude and the parameters governing its short-term plasticity. We used individual pairs of pre- and postsynaptic spike trains to compare the parameters of short-term plasticity, estimated by the models, to the ground truth values.

Simulation: Leaky integrate-and-fire model with adaptation.

To examine the limitations of our models more thoroughly, we simulated a leaky integrate-and-fire model neuron receiving presynaptic input with short-term synaptic plasticity. In particular, to examine the effect of spike frequency adaptation we simulate a postsynaptic neuron with and without an after-hyperpolarization current [56]:

$$\tau_m \frac{dV_m}{dt} = -(V_m - E_{rest}) - r_m g_{sra}(t)(V_m - E_k) + r_m I(t) \qquad (10)$$

$$\tau_{sra}\frac{dg_{sra}(t)}{dt} = -g_{sra}(t)$$

$$\text{if } V_m = V_{th} \text{ then } \begin{cases} V_m \rightarrow V_{reset} \\ g_{sra}(t) \rightarrow g_{sra}(t) + \Delta g_{sra} \end{cases}$$

where E_k is the reversal potential due to K^+, $g_{sra}(t)$ is the spike rate adaptation conductance, which changes with rate $\Delta g_{sra} = 200nS$, and $E_k = 80\ mV$ is the reversal potential. The other parameters were set to $E_{rest} = 80mV$, $E_k = 80mV$, $V_{reset} = 80mV$, $V_{th} = 54mV$, $\tau_m = 10ms$, $\tau_{sra} = 100ms$, and $r_m = 10M\Omega$. Similar to the in vitro experiment above, $I(t)$ is synthesized by simulating a presynaptic input with short-term synaptic plasticity (inhomogeneous Poisson spiking with Tsodyks-Markram PSC amplitudes). We then adjust the DC current, noise, and synaptic strength to achieve the desired postsynaptic spike rate (5Hz) along with a cross-correlogram similar to those obtained by the strongest synapses in the in vitro experiment.

Simulation: Stochastic model of short-term synaptic plasticity

Although the TM model treats short-term synaptic plasticity as a deterministic process, synaptic transmission is a discrete, stochastic process where a discrete number of vesicles are present and probabilistically released following presynaptic spikes. To model this additional variability, we use LIF simulations, as above, where rather than having PSC amplitudes be synthesized from the TM model we use a quantal, stochastic extension of the TM model.

First, to make the TM model discrete, we consider an integer number of release sites, n_{max} , where, at any point in time, only a fraction of resources are available to be released, $a_m = \lfloor n_{max}\ R_m \rfloor$. Following each presynaptic spike, a discrete number of vesicles is released

$$k_m \sim Binomial(a_m, u_m) \qquad (11)$$

giving the PSC amplitudes

$$I_m = \frac{k_m}{n_{max}} \tag{12}$$

Following a spike, the resources and utilization at the following spike (after interval Δt) are given by

$$\begin{aligned} R_{m+1} &= 1 - \left(1 - \frac{a_m - k_m}{n_{max}}\right) e^{-\frac{\Delta t}{D}} \\ u_{m+1} &= U + (u_m + f(1 - u_m) - U) e^{-\frac{\Delta t}{F}} \end{aligned} \tag{13}$$

Results

Here we develop two model-based approaches to estimate short-term plasticity (STP) from trains of pre- and postsynaptic spikes. Both approaches are based on a generalized linear model (GLM) that predicts postsynaptic spiking as a function of the recent history of presynaptic spikes and the postsynaptic spikes. In the conventional GLM, the effect of presynaptic spikes is constant. In the new models we introduce a time-varying coupling term that depends on the history of presynaptic spikes and captures the short-term plasticity of synaptic connections.

In the first model, the coupling term is assumed to vary according to a Tsodyks-Markram model (Fig. 2.1, TM-GLM). The TM model provides a comprehensive description of STP using 4 physiologically motivated parameters: the baseline utilization of resources (U), the magnitude and time constant of facilitation (f and F), and the time constant for the recovery of resources (D). The dynamics of the synaptic resources and their utilization are described by two coupled differential equations that determine how postsynaptic responses depend on the history of presynaptic activity (Eq. 3 in the Methods). Using pre- and postsynaptic spike trains, the TM-GLM estimates both traditional GLM parameters (influence of postsynaptic spiking and coupling between pre- and postsynaptic activity) and the parameters $\theta = \{D, F, U, f\}$ describing short-term plasticity in the TM model.

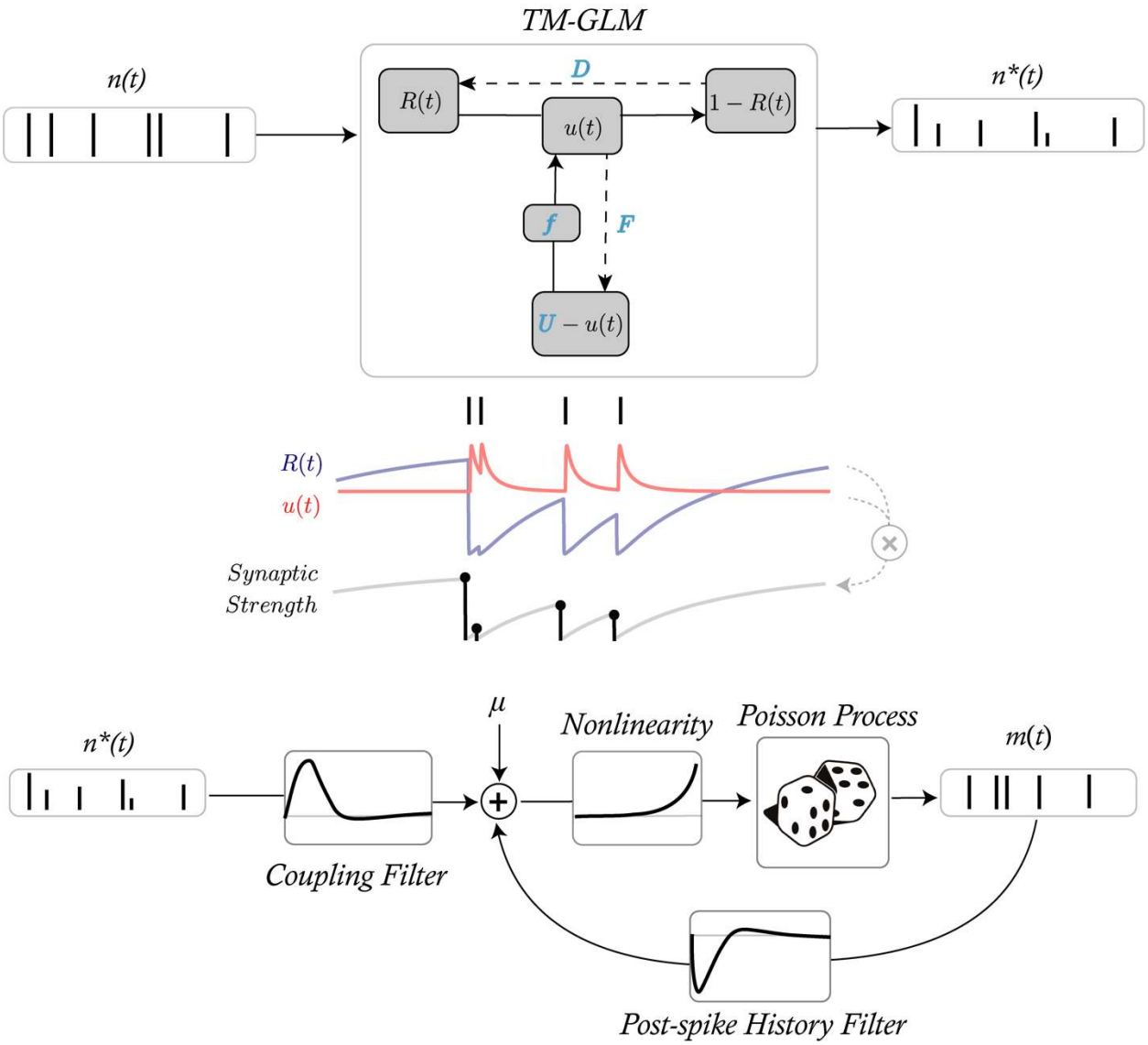

Fig. 2.1: TM-GLM. Here we use the output of the TM as an input to a generalized linear model (GLM) where the weighted presynaptic spikes n(t) combines with the history of the postsynaptic spiking to predict future postsynaptic spiking activity (bottom). The goal of our TM-GLM framework is then to estimate both the parameters of the TM (D, F, U, and f) and the parameters of the GLM (baseline firing rate, coupling filter, and post-spike filter) given only observations of pre- and postsynaptic spiking.*

In the second model, we implement short-term plasticity as a descriptive rule which modifies the coupling term of the GLM based on specific presynaptic inter-spike intervals (ISIs). In this generalized bilinear model (GBLM, Fig. 2.2) the modification rule of the coupling term is not constrained by the known presynaptic mechanisms of short-term plasticity at unitary connections. However, the GBLM can still distinguish between facilitation (where presynaptic spikes following short ISIs have larger postsynaptic effects) and depression (where spikes following long ISIs have larger effects).

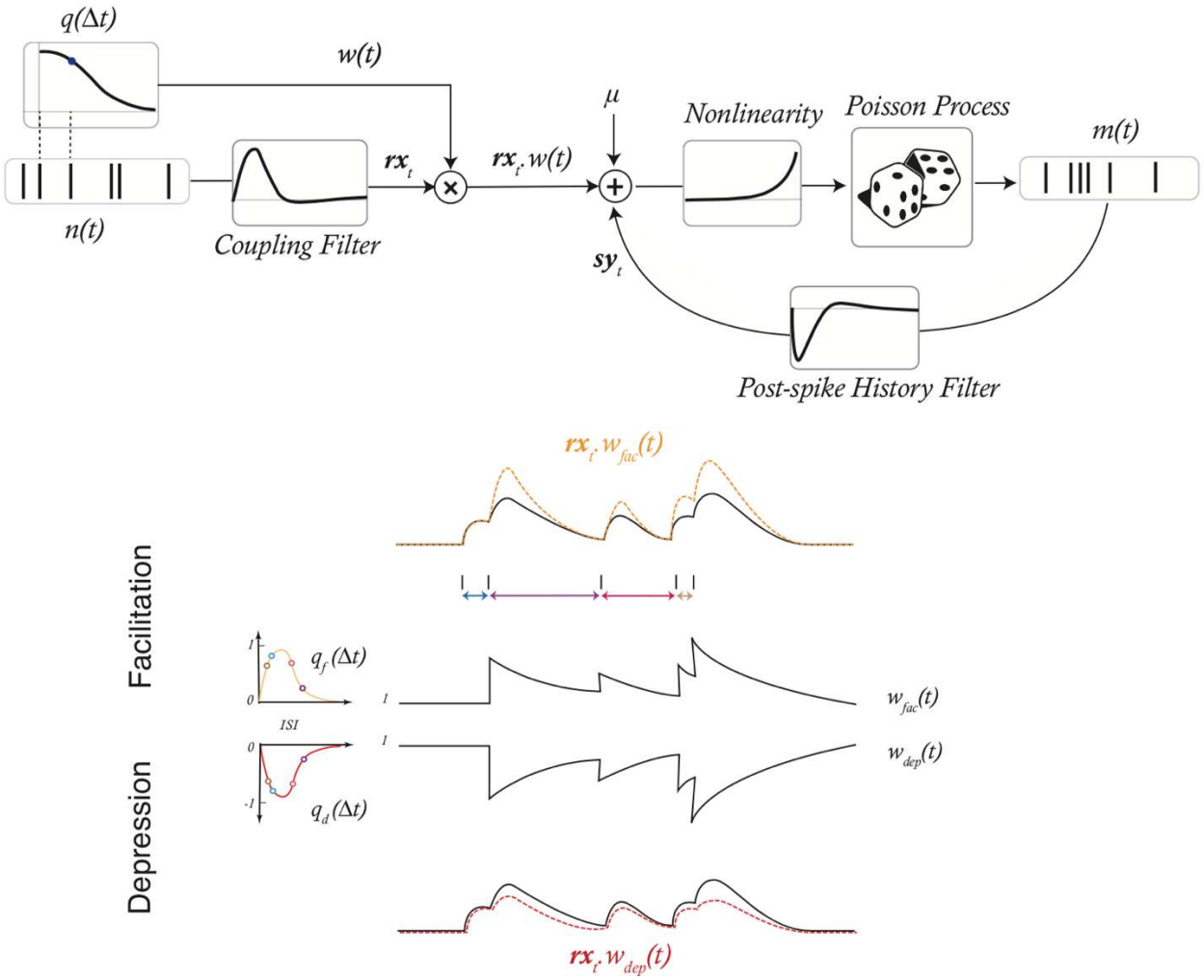

Fig. 2.1: A generalized bilinear model (GBLM) can provide a descriptive model of how synaptic weight varies as a function of presynaptic inter-spike intervals. Here we again use a GLM with coupling from a pre- to postsynaptic neuron and post-spike history dynamics. However, the coupling term is weighted by a separate synaptic weight time-series w(t). Here we assume that w(t) is updated following presynaptic spikes according to a modification function $q(\Delta t)$ and decays exponentially. When the modification function is positive the synaptic weight has facilitating dynamics, and when the modification function is negative the synaptic weight has depressing dynamics (bottom). The goal of our GBLM framework is to estimate the modification

function (q) and the parameters of the GLM (coupling filter and post-spike filter) given only observations of pre- and postsynaptic spiking.

To validate the models, we examine how accurately they can reconstruct synaptic dynamics from spike trains of pairs of neurons connected by synapses with known plasticity rules. We obtained such data using the spiking of layer 2/3 pyramidal neurons evoked by injection of a fully-defined current generated by a population of simulated presynaptic inputs [55]. The advantage of these data is that they are generated by real neurons, with physiological spike generation mechanisms and post-synaptic dynamics. To examine the possible effects of additional factors that are present in *in vivo* recordings and may affect estimation of STP, we also used spike trains generated by simulated leaky integrate-and-fire neurons with: 1) spike frequency adaptation, 2) stochastic release at synaptic inputs, 3) spike sorting errors, and 4) correlated common input.

Current Injection Experiments with Known Short-Term Synaptic Plasticity

To mimic recordings from pairs of neurons with known connectivity and short-term plasticity we made intracellular recordings from layer 2/3 pyramidal neurons in slices of rat visual cortex, and recorded spiking responses of neurons to injection of fully-defined fluctuating current. The injected current was designed to mimic the postsynaptic effect of synaptic inputs which have different strength and express unique synaptic dynamics (Fig. 2.3). To synthesize the current, we used a population of 96 presynaptic neurons where the spike times of each neuron were generated using an inhomogeneous Poisson process with a mean rate of 5 Hz. Six pools of 16 neurons (8 excitatory and 8 inhibitory in each pool) expressed five distinct types of STP, each defined by a unique set of parameters and ranging from strong depression to strong facilitation, along with a sixth pool of neurons which did not express STP (Table 1). STP of synaptic responses was

implemented according to the TM model. Average synaptic weights for the 16 inputs in each pool ranged from strongly excitatory to strongly inhibitory, with excitatory and inhibitory inputs having the same amplitudes but opposite signs. This resulted in a balanced fluctuating current.

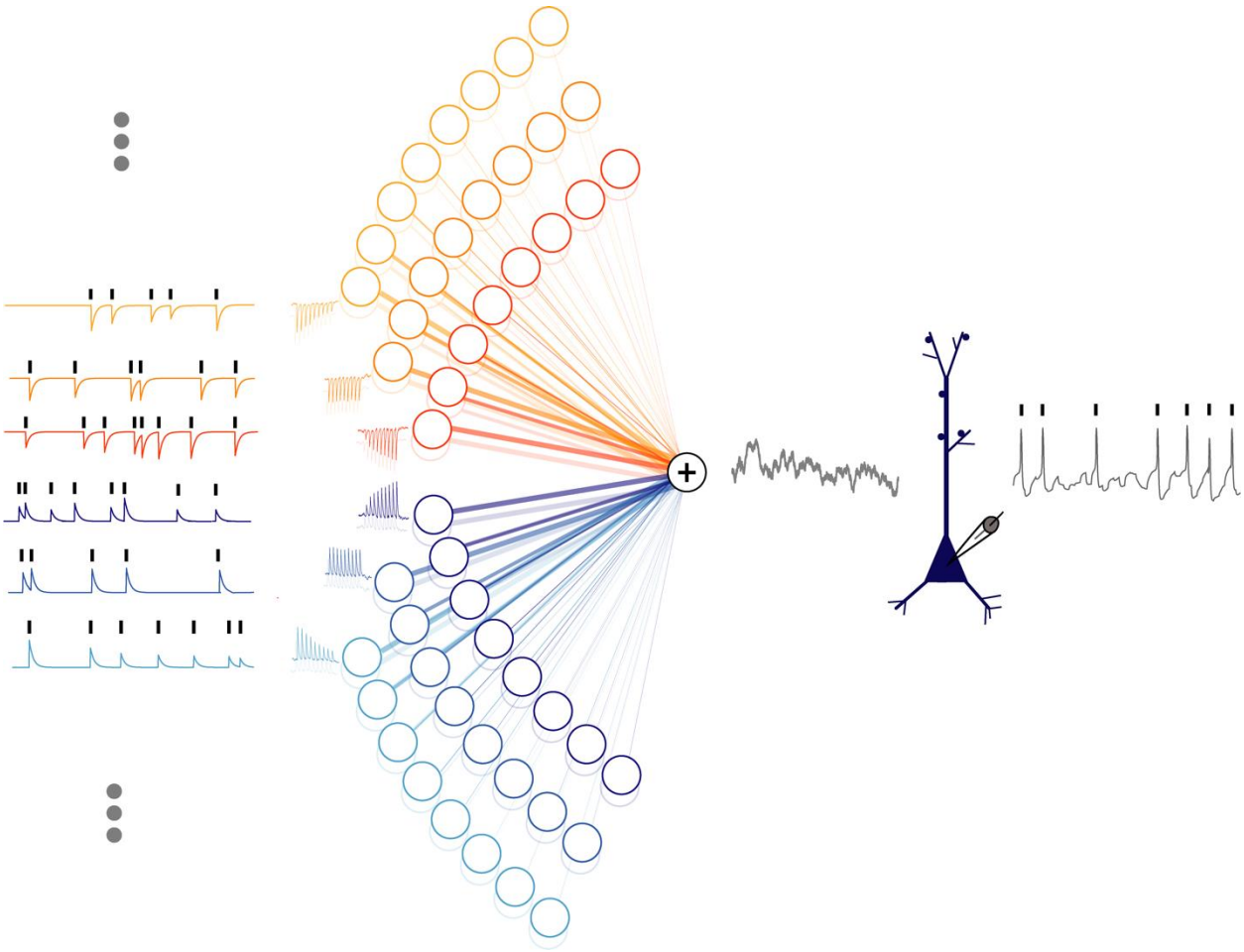

Fig. 2.2: Artificial current injection to a Layer 2/3 pyramidal neuron. To validate our models, we recorded intracellularly from a cortical slice. We first simulated the spike times of 96 presynaptic neurons, then generated postsynaptic current traces corresponding to each input. Inputs had different types of plasticity ranging from strong depression to strong facilitation and a range of synaptic weights (both inhibitory and excitatory). The 96 current traces were then summed together and were intracellularly injected into the postsynaptic neuron whose spiking activity was recorded. These data then allow us to examine the relationship between pre- and postsynaptic spiking under 96 different plasticity/weight conditions.

Table 1: The five parameter sets used to simulate presynaptic currents.

TM parameters	D(s)	F(s)	U	f
Strong Depression	1.70	0.02	0.70	0.05
Depression	0.50	0.05	0.50	0.05
Facilitation-Depression	0.20	0.20	0.25	0.30
Facilitation	0.05	0.50	0.15	0.15
Strong Facilitation	0.02	1.00	0.10	0.11

Using the membrane potential responses to the injected current we detected postsynaptic spikes as positive-slope zero crossings. Thus, in this dataset we knew the timing of presynaptic spikes of each simulated presynaptic neuron, the time-varying synaptic weight, and the timing of the postsynaptic spikes.

To illustrate how STP at a single synapse affects postsynaptic firing in the presence of many other inputs, we performed a separate recording where the injected current had additional structure. One out of 96 presynaptic neurons repeatedly discharged with a pattern typically used for testing STP in slice experiments (9 regularly spaced spikes + 1 after a delay), while the spiking of the remaining 95 presynaptic neurons followed uncorrelated inhomogeneous Poisson processes as described above. This resulted in a repeating test pattern at one synapse embedded in fluctuating noise produced by the activity of the remaining presynaptic neurons. The strength of this synapse was increased to increase signal-to-noise ratio. The average postsynaptic current, membrane potential, and peristimulus time histogram of spiking (PSTH) in response to the test stimulation patterns demonstrate that the effect of a single strong input (>100pA) is clearly observable [Fig. 2.4]. Moreover, synapses with different short-term synaptic dynamics: depression, facilitation and no

plasticity produce distinct postsynaptic responses at all levels. In recordings with *in vivo*-like activity, the effects of short-term synaptic plasticity will be more subtle, since presynaptic spike times do not occur in such regular, repeating patterns under natural conditions and synaptic weights in neuronal connections are much weaker. The remaining analysis focuses on the recordings without the test patterns, where the strongest synaptic weights were ~30pA.

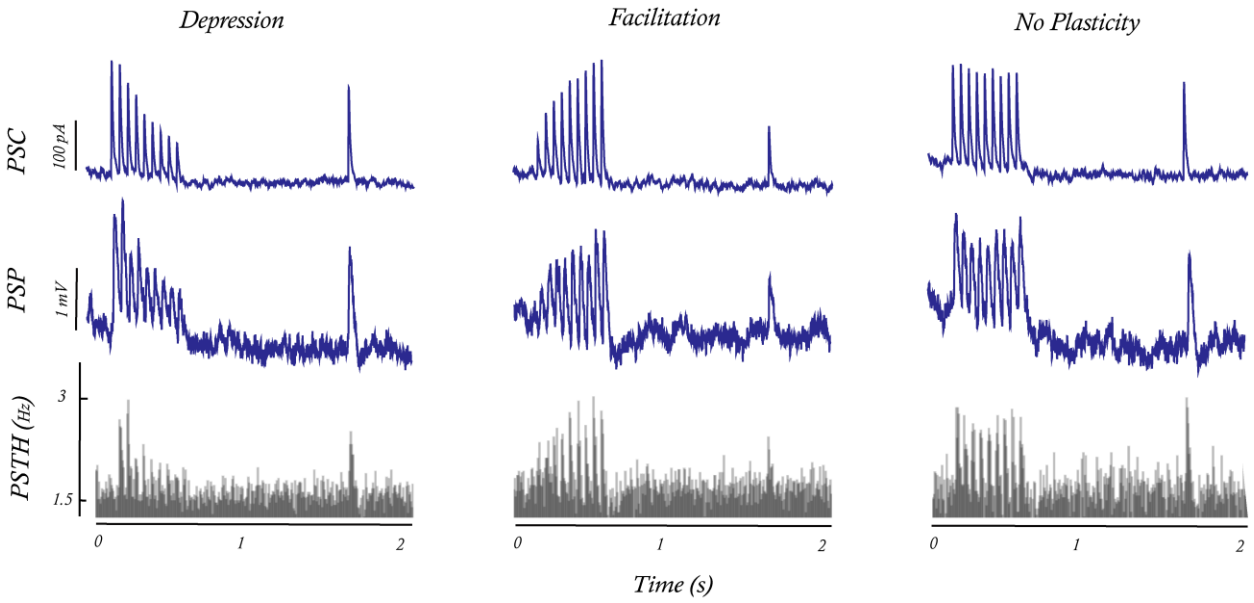

Fig. 2.3: Postsynaptic responses to injection of a regular spike pattern immersed in fluctuating noise in vitro. To verify that the distinct types of STP directly affect postsynaptic spike statistics we compare the responses to a 50Hz train of presynaptic spikes for simulated synapses with short-term synaptic depression (left), facilitation (middle), and without plasticity (right). Only one out of 96 simulated presynaptic neurons had this regular activity pattern; the other 95 presynaptic neurons generated Poisson sequences of spikes to mimic the in vivo setting where postsynaptic neurons receive many presynaptic inputs (SNR=-17dB). Average of n=3000 repetitions shows clear effects of plasticity in the postsynaptic current and potential, but also in the postsynaptic spiking (PSTH).

In previous studies a split-correlogram approach had been used to reveal the effects of short-term plasticity on postsynaptic spike statistics *in vivo* [20,44]. By calculating cross-correlograms separately for presynaptic spikes following short ISIs (or in bursts) and for spikes following long ISIs (isolated spikes), evidence was found for both short-term facilitation [44] and depression [20,21] of synaptic transmission *in vivo*. To determine if this method of analysis could reveal

effects of STP in our data obtained with inhomogeneous Poisson presynaptic spiking, we split presynaptic spike trains into spikes following inter-spike intervals shorter than the 10th percentile and longer than the 90th percentile of ISI distribution (Fig. 2.5A). Separate analysis of the postsynaptic effects of presynaptic spikes from these two groups revealed clear differences between synaptic inputs with distinct types of plasticity [Fig. 2.5B]. In connections with depressing synapses the PSCs, PSPs, and, most importantly, peak spike counts in the correlograms were much reduced for short intervals. In connections with facilitating synapses, the postsynaptic effects were slightly increased following short presynaptic intervals [Fig. 2.5B]. In synapses with intermediate forms of plasticity the effect of ISI on postsynaptic responses was less pronounced and was between the two extremes. Note that because of temporal summation after short ISIs the increase of the postsynaptic responses (PSCs, PSP, and spike count) is evident in the short interval correlograms even shortly before 0 ms, similar to the results from *in vivo* study [44].

Thus, the effects of STP on spike responses of neurons to injection of a fully-defined current were clearly expressed in the difference between split-correlograms, consistent with results reported for *in vivo* recordings [20,21,44]. Our results show that the effects of ISI on split-correlograms were more pronounced for depressing than for facilitating inputs. One possible reason for such asymmetry may be that the presynaptic spike statistics used here does not fully elicit the effects of facilitation. To address this issue, we examined the distribution of PSP amplitudes as a function of inter-spike intervals for synapses with the different types of STP used in our model [Fig. 2.5C]. While in depressing synapses the PSP amplitudes monotonically increase as ISIs increase, the response amplitudes in facilitating synapses depend on the ISIs in a non-monotonic way. At facilitating synapses, there is an ISI range in which PSP amplitudes are elevated, but for both shorter and longer ISIs the amplitudes are reduced (Fig. 2.5C). This pattern makes it difficult to

distinguish facilitation in split cross-correlograms, since short and long ISIs can produce similar PSP amplitudes. Moreover, facilitating responses also have higher variability than depressing responses for any given ISI, likely since stronger facilitation enhances the variability of utilization (release probability) compared with depressing synapses (Eq. 3). These factors appear to hinder detection of short-term facilitation with split-correlogram analyses.

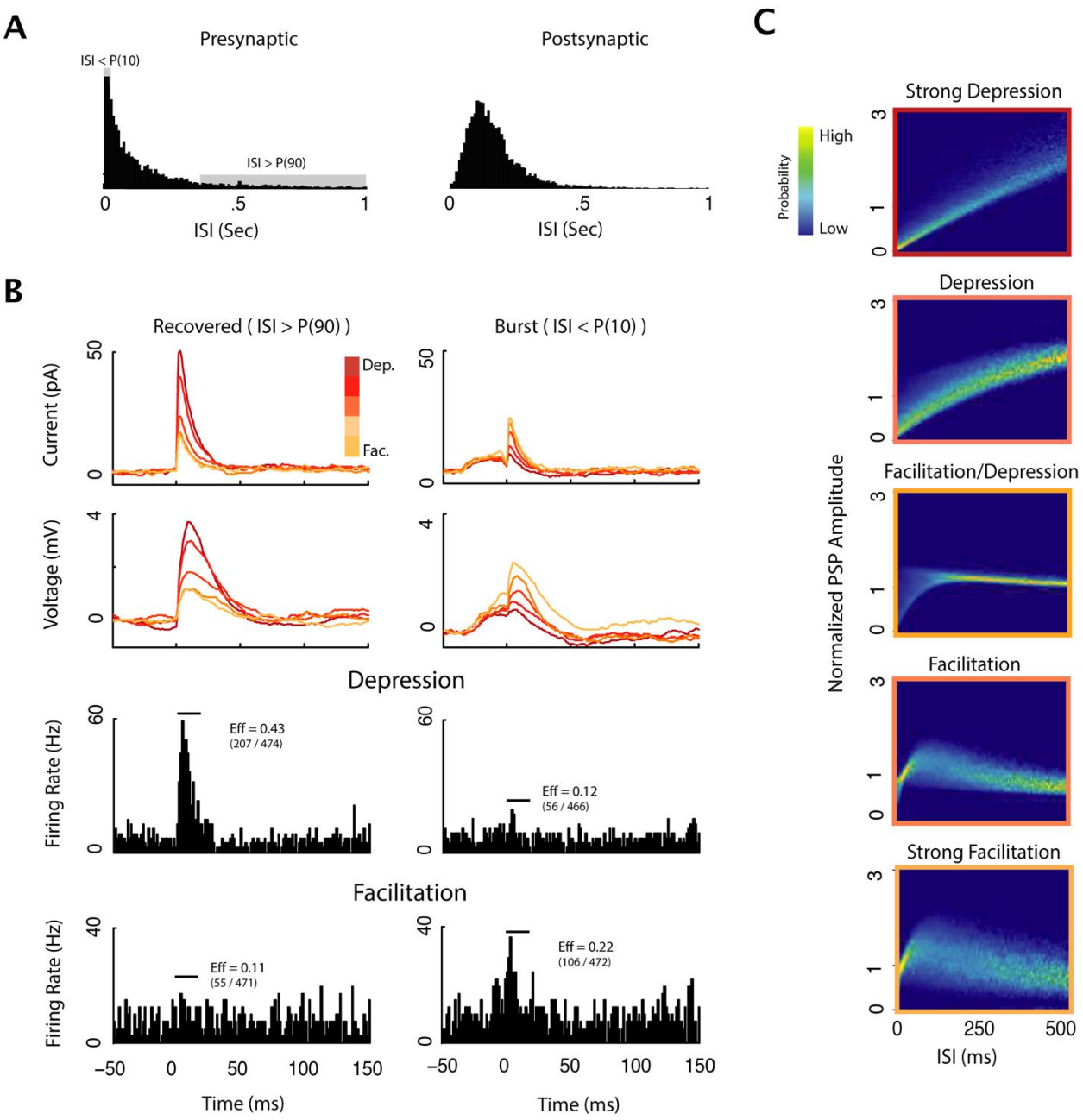

Fig. 2.4: Postsynaptic responses for different presynaptic ISIs reveal the effects of STP. A) Inter-spike interval (ISI) distribution of a presynaptic and the postsynaptic neuron B) PSCs and PSPs [top] for strongest excitatory synapses for five different types of plasticity (colors) separated by presynaptic ISI (<10th percentile, left; >90th percentile, right). Depressing synapses have much larger PSC/PSPs for ISIs >90th percentile, while facilitating synapses have larger PSC/PSPs for

the ISIs <10th percentile. Postsynaptic spiking [bottom] shows similar effects, but tend to be much more difficult to interpret due to the sparse spike responses. Split cross-correlograms are shown for two synapses: one with strong depression and one with strong facilitation [bottom]. For comparison, the PSPs and PSCs are vertically offset such that the average from -50ms to -30ms was set to 0. C) The distribution of PSP amplitudes as a function of presynaptic ISI for the different classes of plasticity used in experiment. One reason that split cross-correlograms are difficult to interpret is that there is no deterministic relationship between ISI and PSP amplitude, and, in some cases, such as with facilitating synapses, the relationship is nonlinear.

The examples considered above show results for the strongest, excitatory simulated inputs (~30 pA). Weaker excitatory synapses and inhibitory synapses express similar dynamics in their PSC and PSP amplitudes, however, the postsynaptic effects are less pronounced and show greater variability. For weak facilitating synapses there is often no detectable difference between the postsynaptic responses to short and long intervals. This analysis exposes a fundamental drawback of the split-correlogram approach: its low sensitivity to transient effects. By explicitly modeling how synapses vary in response to the history of presynaptic spiking, rather than modeling the average responses to only a single previous ISI, model-based approaches can more accurately reconstruct synaptic dynamics and distinguish between different types of STP.

Inferring STP parameters from spike trains using the TM-GLM

We extend the GLM framework to include short-term synaptic plasticity implemented according to the Tsodyks-Markram model (see Methods). The TM model describes the dynamics of synaptic transmission using two coupled differential equations for resources R and their utilization (release probability) u with a set of four parameters $\boldsymbol{\theta} = \{\boldsymbol{D}, \boldsymbol{F}, \boldsymbol{U}, \boldsymbol{f}\}$ (Eq. 3 in the Methods). To fit the

TM-GLM to the observed spike trains we use an alternating coordinate ascent to maximize the (penalized) likelihood of observed postsynaptic spiking. Namely, we update the plasticity parameters with fixed GLM parameters and then update the GLM parameters with fixed plasticity parameters, alternating between the two optimization problems until the maximum is achieved. The TM formalism assigns a weight to each spike of the presynaptic neuron, while the GLM parameters characterize the influence of prior postsynaptic spiking and coupling between pre- and postsynaptic activity (as scaled by the TM weights). To facilitate convergence of the TM and the GLM parameters we impose prior constraints on both these parts of the model (see Methods). Using pre- and postsynaptic spike trains, we thus obtain estimates of both traditional GLM parameters and a complete set of parameters $\boldsymbol{\theta} = \{\boldsymbol{D}, \boldsymbol{F}, \boldsymbol{U}, \boldsymbol{f}\}$ describing short-term plasticity in the TM model.

We fit the TM-GLM separately for each simulated connection in our in vitro recording. The 96 simulated presynaptic inputs had different weights and different types of STP, and our goal is to compare how these synaptic properties affect estimation of STP. Specifically, we have six sets of parameters corresponding to strong depression, depression, depression/facilitation, facilitation, strong facilitation, and a control set with no plasticity (Table 1). Although the optimization of the TM parameters is not convex, we find that, after adding informative priors (see Methods) the global optimum can be quickly found using random restarts. TM-GLM estimates of the time constant for depression D and the release probability U are closer to underlying true values than the estimates of the facilitation time constant F and its magnitude f. Fig. 2.6A shows results of bootstrapping to estimate the parameter uncertainty for the different types of plasticity. Note that high variability in the estimation of facilitation parameters is not a specific drawback of our model, but represents a more general problem. Indeed, previous work showed that estimates of facilitation

parameters were non-precise even when direct measurements of postsynaptic responses, PSPs or PSCs (and not postsynaptic spikes as used in our model) were fitted [33,37]. Particularly for depressing synapses (where U is large and F is small), the estimation of f is not well-posed. In this case, it may make more sense to use a more restricted TM model with fewer parameters [11,32] or to use a fully Bayesian approach where the posterior can be more completely assessed. More generally, the difficulty of estimating facilitation parameters might be a consequence of a relatively weaker effect of facilitation on postsynaptic activity as compared to depression. This interpretation is supported by the observation that despite the deviation of estimated parameters of facilitation from the true value, the model with the estimated parameters accurately predicts the steady-state filtering properties of dynamic synapses (Fig. 2.6C), as well as split cross-correlogram (Fig. 2.8A). Note that some of the bias in parameter estimation may be due to the choice of priors. Here we chose our priors to avoid local minima in the posterior that occur near the edges of parameter space, where F or f are close to zero. However, as the number of observations increases these biases will be reduced, since likelihood will have a larger impact on the posterior than the prior. In general, the accuracy and confidence of the estimates will be affected by many factors, such as, the number and pattern of presynaptic spikes, number of postsynaptic spikes, the synaptic weight, and the type of STP.

For large-scale analysis of STP in neuronal networks it might be important to distinguish between different types of plasticity at a synapse (e.g. facilitating vs depressing) and attribute certain types of plasticity to different classes of synaptic connections, rather than to extract the exact parameter values for each synapse. Again, although the problem is not convex, we find that the different types of plasticity can be distinguished based on spiking observations alone. For the 5 strongest excitatory inputs with each type of plasticity we compare the likelihood under the different settings

of the TM parameters used in the recording [Fig. 2.6B]. This analysis treats the problem of STP-identification as a classification problem. If the data do not provide a clear indication of the type of STP, e.g. for very weak synaptic inputs which have little effect on postsynaptic spiking, then the likelihood should be similar under all models – both facilitating and depressing. However, here we find that the true parameters do have the highest likelihoods, with depressing inputs having high likelihoods under the depressing model and facilitating inputs have high likelihoods under the facilitating model.

Additionally, even though the estimated parameters may differ from their true values, the (steady-state) synaptic dynamics of the estimated models typically matches the dynamics of the true models [Fig. 2.6C]. Depressing synapses show characteristic low-pass filtering, while facilitating synapses have band-pass filtering with cutoff frequencies depending on the exact TM parameters.

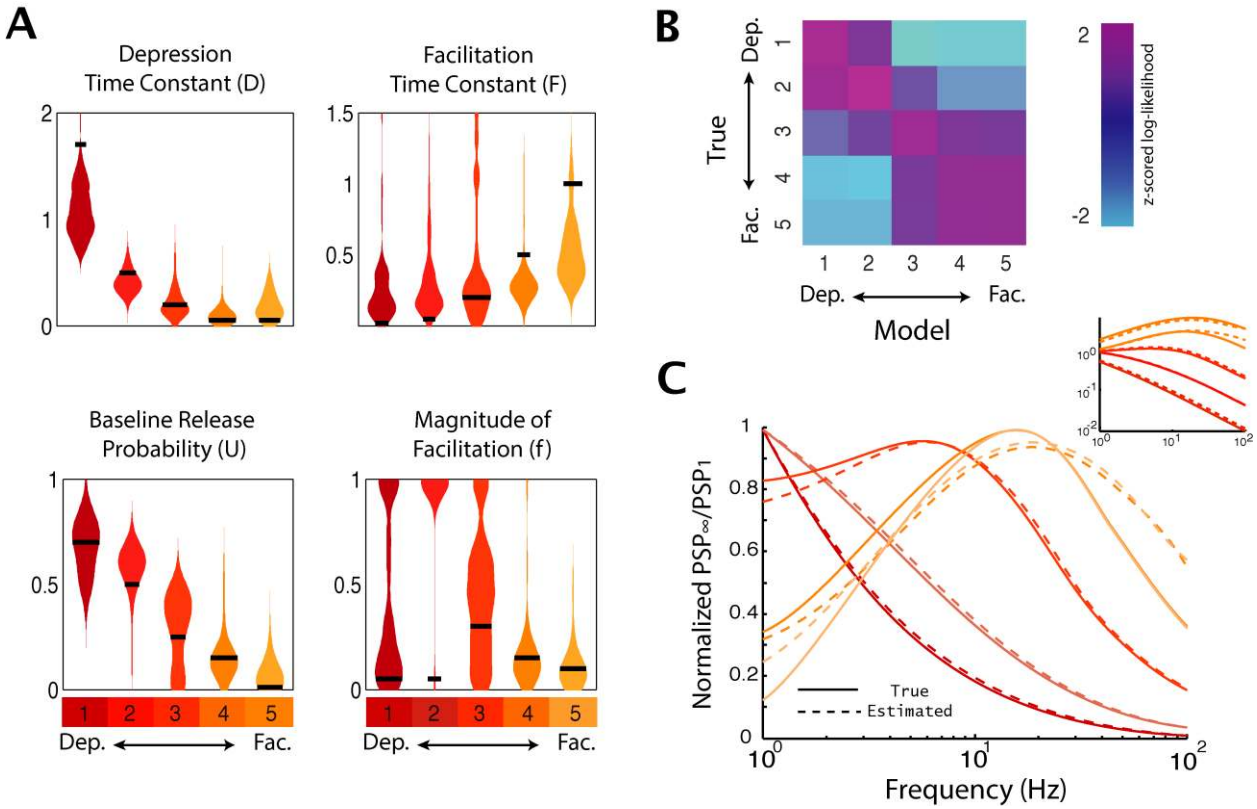

Fig. 2.6: Estimation of the TM parameters. A) Bootstrap distribution of the estimated parameters $\{D, F, U, f\}$ *for synapses with five different types of STP, from strong depression to strong facilitation (left to right). Here only the results for the strongest excitatory synapses are shown. The black horizontal bar in each distribution represents the true value for each artificial synapse (Table 1). B) z-scored log-likelihood values for strong synapses modeled with the parameters of each possible model. Even though the optimization of the TM-GLM is not guaranteed to be convex, we can accurately discriminate between different types of STP. Note that the highest likelihood is along the diagonal where the true type of STP corresponds to the same type modeled. C) Normalized steady-state postsynaptic potentials in response to a regular train of presynaptic spikes with different input frequency for true parameter sets (solid) and estimated parameters (dashed). Inset shows the unnormalized steady-state response on a log-scale.*

Inferring STP from spikes using a Generalized Bilinear Model

The TM-GLM estimates the short-term dynamics of a synapse described with biophysically realistic parameters that are related to the vesicle and calcium dynamics. In many cases, however, it might be useful to detach the description of the coupling between pre- and postsynaptic spiking from the biophysics of synaptic dynamics at an individual synapse. To describe neuronal interactions in terms of ISI-dependent modifications, we introduce a generalized bilinear model (GBLM, Fig. 2.2) that captures functional changes in the synaptic efficacy for different presynaptic intervals. In this model, the coupling term changes as function of presynaptic spiking, e.g. at facilitating synapses it increases for short ISIs, and at depressing synapses it decreases for short ISIs. We use basis splines to fit a smooth modification function (see Methods) that describes how the coupling term has been adjusted following different presynaptic intervals. We further assume that the effect of the modification is transient, decaying exponentially [Fig. 2.2]. Compared to the TM-GLM, the GBLM has simplified description of the dynamics of coupling but provides a more explicit characterization of the effects of different ISIs on the modification of the coupling term.

The GBLM provides clearly distinct estimates of the modification functions for synaptic connections with different types of short-term plasticity [Fig. 2.7]. For simulated inputs expressing the same type of STP, but having different weights (among strongest 3) or different signs (excitatory and inhibitory), the estimates of the modification functions were similar. These modification functions were estimated by maximizing the regularized log-likelihood. For stability, the spline basis was designed to have no effect on very short or very long ISIs where there is typically little data. However, for depressing synapses the modification function decreases the relative synaptic strength for ISIs between 0 and 1s, and for facilitating synapses the modification function increases the relative synaptic strengths.

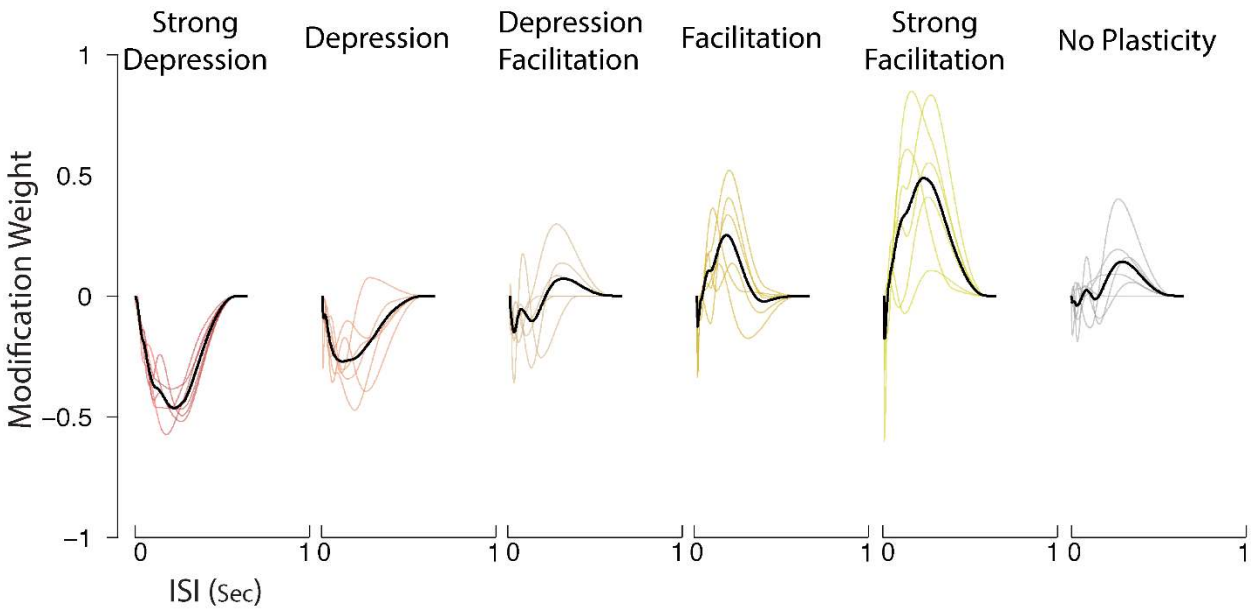

Fig. 2.7: The modification function estimated using the generalized bilinear model (GBLM) for five different types of STP and for synapses with no plasticity. The modification functions for six strongest synapses (three inhibitory and three excitatory) are shown in color and the average is shown in black. These functions describe how synaptic weights change following different inter-spike intervals and allow the different types of STP to be distinguished. For strong depression, the modification function is negative and for strong facilitation the function is positive, capturing the respective decreases and increases in synaptic strengths.

Comparison of the Models

Both the TM-GLM and the GBLM accurately describe split cross-correlograms for all examined types of STP, and for both excitatory and inhibitory inputs for the in vitro experiment [Fig. 2.8A]. However, in addition to the spike statistics we can also compare how well the models reconstruct the time-varying individual PSC amplitudes. After estimating the plasticity dynamics for each simulated input using the TM-GLM (n^*_{pre}) and the GBLM ($w(t) \odot n_{pre}$) we then calculate correlations between the true PSC amplitudes and the estimated amplitudes under the two models

[Fig. 2.8B]. We find that the weights of the simulated inputs have a substantial effect on the reconstruction of PSC amplitudes. The estimated amplitudes at strong synapses (both excitatory and inhibitory) are reconstructed much more accurately than amplitudes at the weak synapses. Additionally, we find that the PSCs of depressing synapses are much more reliably reconstructed than PSCs of facilitating synapses ($r=0.95\pm0.01$ for synapses with strong depression vs. $r=0.34\pm0.06$ for synapses with strong facilitation). This is consistent with our observation that the PSCs of depressing synapses are more reliably related to ISIs compared to facilitating synapses [Fig. 2.5]. Finally, the TM-GLM model appears to consistently out-perform the GBLM (average correlation for the TM-GLM across all types of plasticity and weights is $r=0.70\pm0.03$ compared to $r=0.52\pm0.03$ for the GBLM).

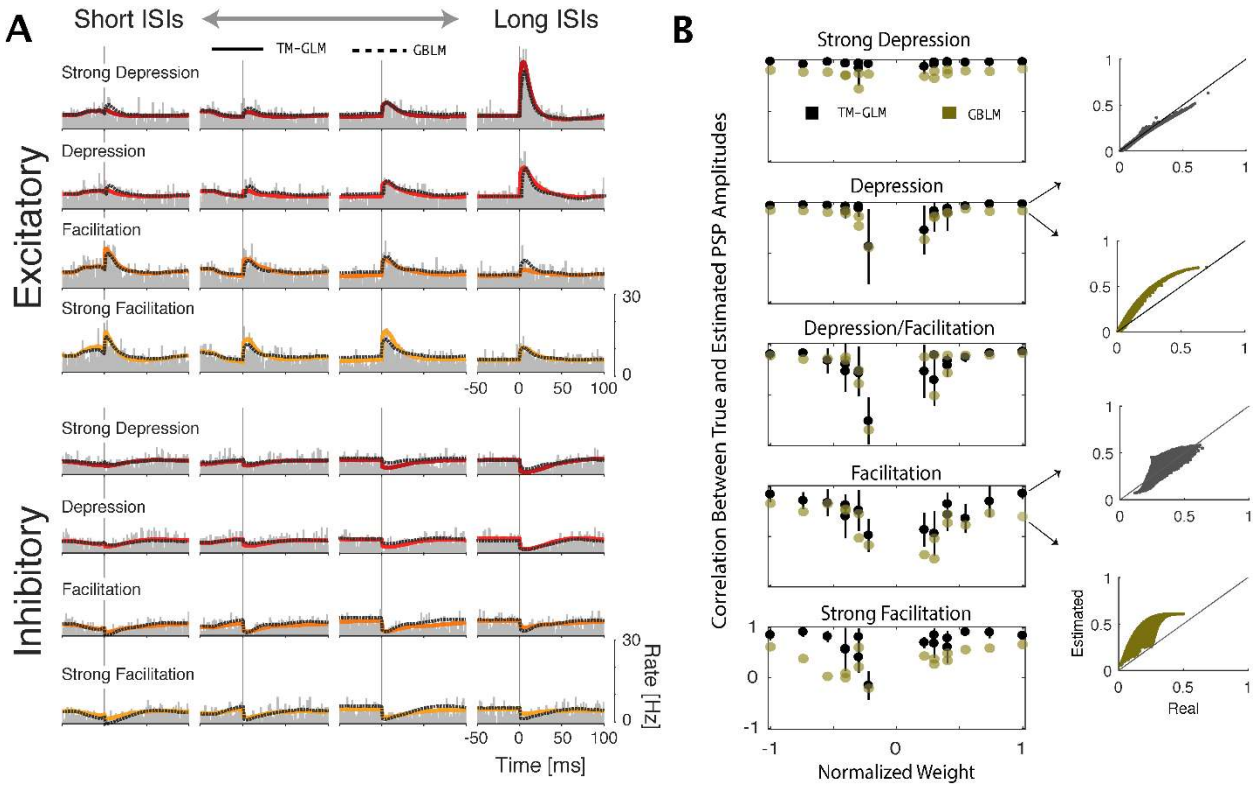

Fig. 2.8: Split cross-correlograms. A) For each type of STP, the cross correlations between pre- and postsynaptic spiking split for the quartiles of the presynaptic ISI distribution (columns) from shorter (left) to longer inter-spike intervals (right). The estimated cross-correlation from TM-GLM (solid) and GBLM (dashed) are shown on top of the observed cross-correlation (gray bars). B) The correlation between true and estimated amplitudes of postsynaptic potentials in five different classes of plasticity as a function of the overall synaptic (based on 1000s recording time with 5Hz presynaptic firing rate). The PSPs of depressing synapses tend to be more accurately reconstructed than those of facilitating synapses, and weaker synapses (both excitatory and inhibitory) tend to be less accurately reconstructed than strong synapses.

Potential problems in raw spike data that may confound estimation of STP

In vitro recordings of responses to simulated presynaptic spikes have the advantage that the postsynaptic spikes are generated by the biophysics of a real neuron. However, estimation of STP from spike trains recorded in the intact brain *in vivo* may be compromised by several additional factors, not considered in this controlled experimental setting. Below we will analyze possible effects of four such factors on STP estimation: spike frequency adaptation, stochastic release of transmitter, uncertainty of spike sorting, and correlated common input. To examine how these sources of variability may affect the estimation of short-term synaptic plasticity from spikes we simulated postsynaptic spike trains using leaky integrate-and-fire model neurons receiving synaptic inputs with defined STP in the presence of noise. For simplicity, we focus on model synapses with strong depression, strong facilitation, and no plasticity (Table 1).

Spike Frequency Adaptation

One factor that affects postsynaptic firing is spike frequency adaptation. In particular, an after-hyperpolarization (AHP) current mediating fast spike frequency adaptation can change the pattern of postsynaptic firing and may act to mask the influence of presynaptic STP on generation of postsynaptic spikes. To test if our models can differentiate the effects of AHP currents (I_{AHP}), which alter the dynamics of the postsynaptic neuron, from the effects of short-term synaptic plasticity, we simulated two leaky integrate-and-fire (LIF) neurons with and without an I_{AHP} [56] (see Methods). In response to a long depolarizing pulse, the LIF neuron without an I_{AHP} fires at a stationary rate. The LIF neuron with the I_{AHP}, on the other hand, rapidly adapts – with a firing rate peaking immediately after the depolarization onset and gradually decreasing to a lower steady-state. After stimulus offset the firing rate of the adapting LIF decreases below the pre-stimulus

level [Fig. 2.9A]. These effects are not due to synaptic dynamics but reflect the dynamics of the postsynaptic neuron itself.

We simulated pre- and postsynaptic spike trains using the LIF model neurons (with and without an I_{AHP}) receiving inhomogeneous Poisson input with short-term synaptic dynamics governed by the TM model and applied our models to estimate STP from these spike trains. Results from the TM-GLM and GBLM for the two leaky IF neurons show that the adaptation properties mediated by I_{AHP} current are mostly captured in the post-spike history filters [Fig. 2.9B-C]. For connections with depression, facilitation, or no STP, the estimated TM parameters and the modification functions estimated with the GBLM are similar with and without the I_{AHP}. Although frequency adaptation occurs on a similar timescale to short-term synaptic plasticity, the methods here thus seem to be able to distinguish purely postsynaptic dynamics from the time-varying effect of the presynaptic neuron on the postsynaptic neuron.

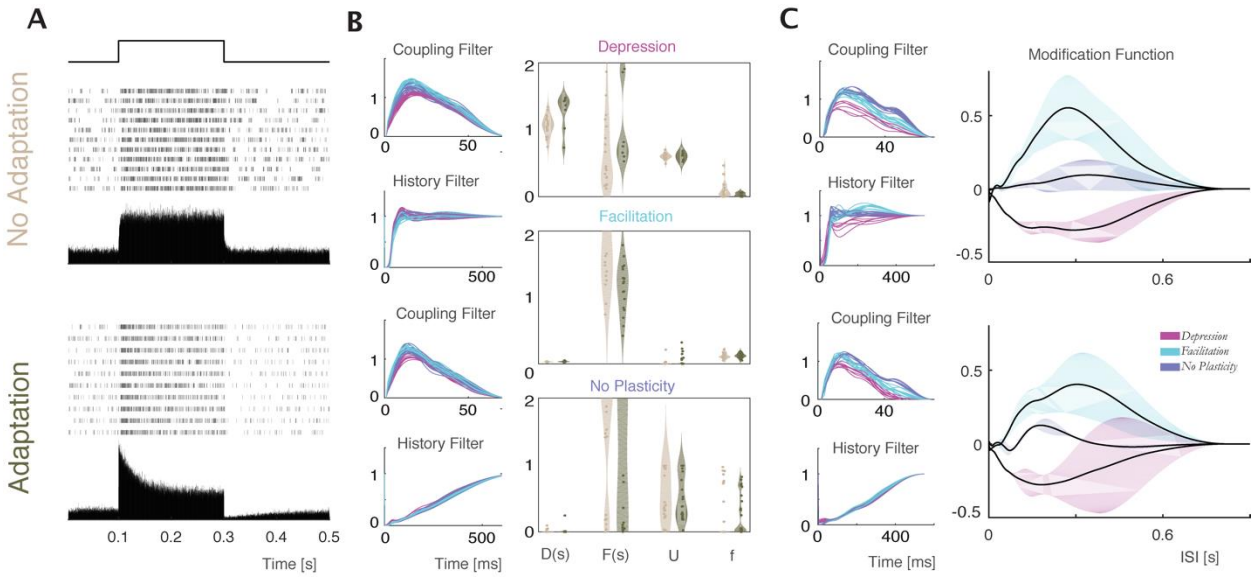

Fig. 2.9: Spike frequency adaptation affects post-spike history filters but does not affect STP estimation. A) Spiking of two LIF model neurons, with and without an I_{AHP} current, in response to a long depolarizing current step. Current step, spike rasters and PSTHs are shown for each model neuron. B) Parameters estimated by the TM-GLM. Estimated coupling filter and post-spike history filter of the two neurons in response to inhomogeneous Poisson inputs with short-time depression (blue), facilitation (red), and no plasticity (turquoise). Violin plots show estimated TM parameters for inputs with each type of the plasticity for the model neurons without (beige) and with I_{AHP} (green). C) GBLM estimates of coupling filter, post-spike history filter, and modification function for depressing (blue), facilitating (red), and no plasticity inputs (turquoise). Solid lines in the right panels show the average modification functions ±1 SD (bands).

Stochastic Release

One further potential source of noise that is not included in the Tsodyks-Markram model, and that was not accounted for in our experiments in slices, is stochastic vesicle release. Although the TM model and the GBLM treat the synaptic transmission as deterministic and the PSC/PSP amplitudes

can take any value, in real synapses PSC/PSP amplitudes are fundamentally stochastic with vesicles being probabilistically released from a limited number of sites. Compared to our in vitro experiments using the deterministic release, it may be more difficult to estimate STP parameters from the spiking of real neurons with stochastic release. To study how stochastic release impacts the estimation of STP parameters, we use a quantal model of synaptic plasticity [53,57]. In this model, the resources of the TM model are discretized based on the number of release sites and are then released according to a Binomial distribution with a time-varying probability given by the utilization variable of the TM model (see Methods). We simulated pre- and postsynaptic spike trains from LIF model neurons driven by inhomogeneous Poisson input with synaptic dynamics governed by the quantal TM model. The amplitudes of the postsynaptic currents are now noisy rather than deterministic functions of the presynaptic spike timing. In our simulations, increasing the number of release sites decreases the variance of the PSC amplitudes. For depressing synapses, stochastic release leads to a systematic bias in the estimates of the TM model parameters compared to their values under deterministic release [Fig. 2.10A]. For facilitating synapses, on the other hand, the TM parameter estimation was not substantially affected. Similarly, the modification functions estimated with GBLM for depressing synapses were changed as the number of release sites is varied, while the modification functions for facilitation are more stable. Both the TM-GLM and GBLM can still distinguish between depression and facilitation, but considering stochastic release may be necessary for accurate parameter estimates *in vivo*.

Spike Sorting

Another potential source of uncertainty, that may affect the estimation of synaptic dynamics from spikes, is imperfect spike sorting. In practice spike sorting from *in vivo* recordings is not a perfect process, and inaccuracies in spike sorting can lead to biased estimates of neural response properties

[58]. Here, we simulated presynaptic and postsynaptic spike trains using LIF model neurons with strongly depressing or facilitating dynamics on inhomogeneous Poisson input (as above, See Table 1 for parameters). We then simulated the effects of imperfect spike sorting by randomly deleting and inserting spikes into both the pre- and postsynaptic spike trains before estimating STP. For insertion, we randomly selected spikes from two other inhomogeneous Poisson neurons (same baseline firing rates) and assigned the spikes to pre- and postsynaptic neurons. For both the TM-GLM and GBLM we find that the imperfect assignment of spikes (both addition and deletion) results in only small biases in the estimation of STP parameters for connections with strong facilitation and depression [Fig. 2.10B]. Despite these small biases, we were able to distinguish between facilitation and depression even as the proportion of spike sorting errors becomes large (20-40% insertion/deletion).

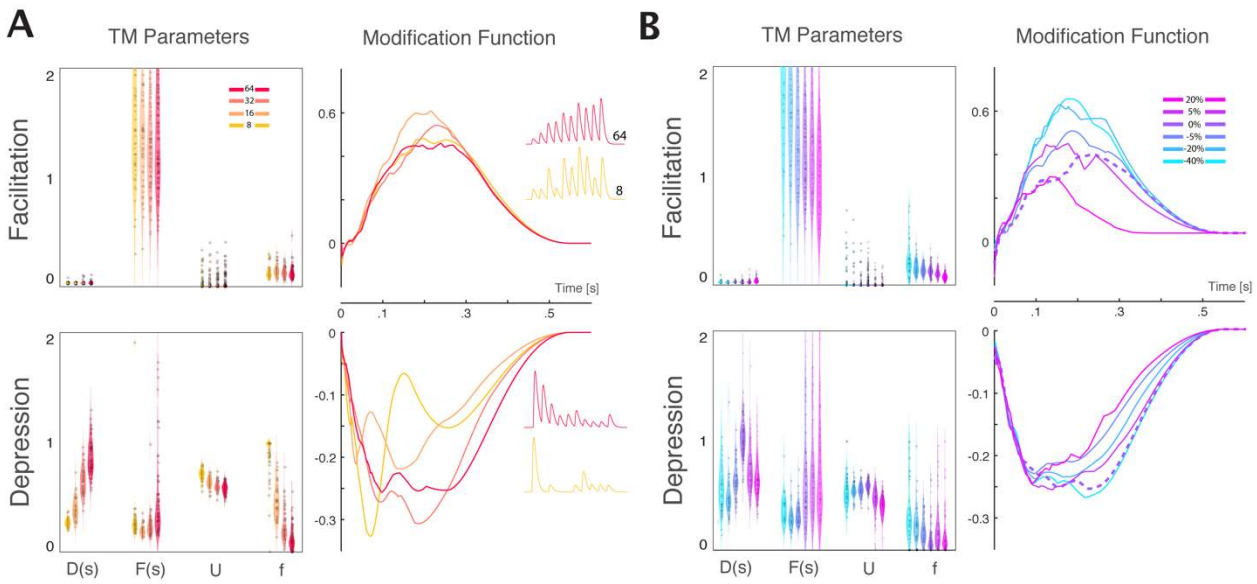

Fig. 2.10: A) Stochastic vesicle release leads to a bias of STP estimates. Left panels; TM parameters for simulations with four different numbers of release sites: 8, 16, 32, and 64 (color coded). Right panels; median modification functions for 4 different release site numbers. Colors match with the left panel. B) Inaccurate spike sorting does not substantially alter STP estimates. Left panels; TM parameters estimated using simulated spiking of LIF neurons with random deletion or addition of spikes. From left to right: random deletion of -40%, -20%, -5% spikes; without deletion or insertion (0%), and with insertion of 5%, and 20% of spikes. Right panels: median modification functions for the same sets with deleted and inserted spikes (colors match with the left panel). Dashed line denotes the estimated modification function under perfect spike sorting (without insertion or deletion).

Common Input

In vivo, neurons often have common synaptic input from unobserved sources. Common input introduces correlations in pre- and postsynaptic spiking that are not due to synaptic connections between the recorded neurons. To study how such correlations would affect STP estimation we simulated a microcircuit with different levels of synchrony. In this simulation, two presynaptic neurons receive input from three sources: 1) a private, slowly fluctuating current, 2) a shared/common, slowly fluctuating current, and 3) an independent white noise current. The postsynaptic neuron receives the common input, an independent white noise current, and inputs from each presynaptic neuron – one with a depressing synapse and one with a facilitating synapse [Fig. 2.11A]. We then vary the strength of the common input using a weight parameter w, which determines how much of each neuron's input is originating from the shared/common source and how much of the input comes from the private current. As the weight of the common input increases there is a short-term synchronization between the spiking of all neurons [Fig. 2.11B]. At low ($w = 0.25$) and medium ($w = 0.5$) common input both the TM-GLM and GBLM were able to discriminate between depressing and facilitating inputs, but at $w = 0.75$ neither model was able to distinguish between the depressing and facilitating input. This simulation demonstrates that, at least in some situations, strong common input can cause both models to fail to estimate underlying short-term synaptic plasticity.

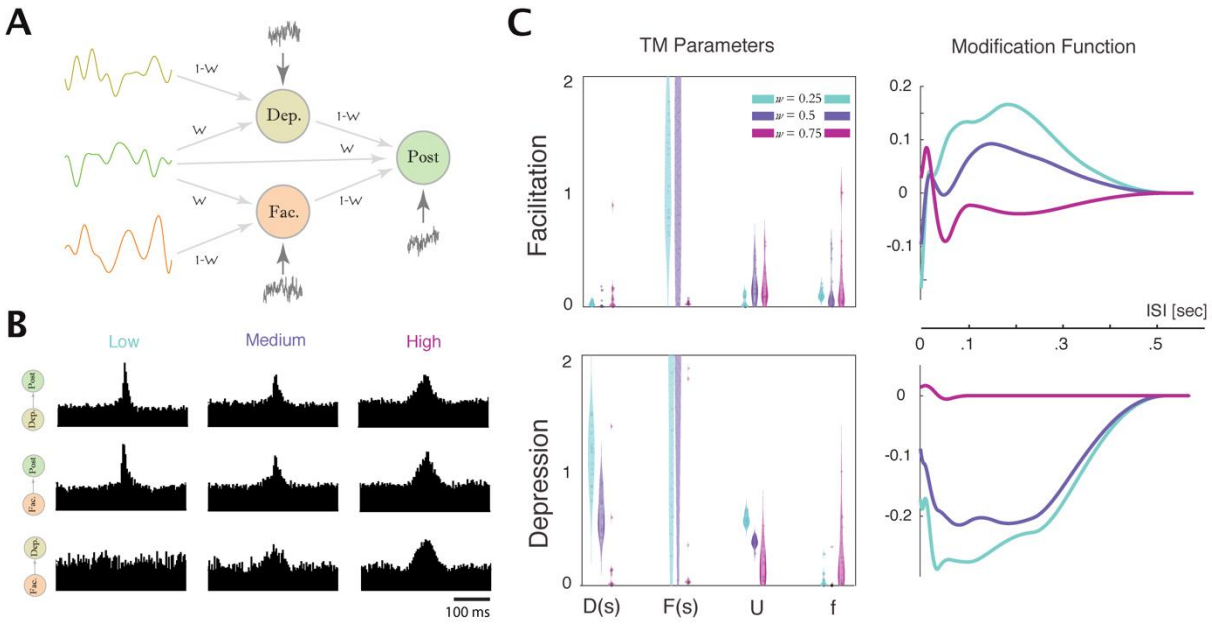

Fig. 2.11: Common input can prevent accurate estimation of STP. A) We simulated a microcircuit where, rather than receiving independent input, the presynaptic and postsynaptic neuron both receive correlated, common input. We use three different sources of fluctuating input to each of the three LIF model neurons, varying the strength of common input. B) Cross-correlograms for three levels of common input (w= 0.25, 0.5, 0.75). C) Left panels: TM parameters estimated from spiking of neurons with (from left to right) low, medium, and high common input. Right panels: GBLM modification functions for the facilitating (top) and depressing (bottom) synapses and the three different levels of common input.

Discussion

Intracellular recordings in brain slices have revealed a diversity of STP across cell types and anatomical connections [5,6]. Moreover, the details of STP at a given type of synapse may change depending on a multitude of factors, such as changes during development [59], neuromodulation [60], or induction of long-term plasticity [31]. Because STP critically affects information processing, understanding operation of neuronal networks during natural behavior requires large-scale analysis of STP *in vivo*. However, since large-scale intracellular recordings are not feasible *in vivo*, alternative methods are necessary for such studies. Large-scale extracellular recordings, on the other hand, are feasible *in vivo*. Existing techniques allow simultaneous recording of spiking of hundreds of neurons, and this number appears to be growing exponentially [17]. Characterizing short-term plasticity using spike observations is more difficult than using intracellular (PSC/PSP) signals, but short-term synaptic plasticity does have observable effects on spike statistics.

Prior evidence for STP *in vivo* obtained from spike trains alone employed a split cross-correlogram approach, in which the postsynaptic response to presynaptic spikes following short ISIs was compared to that following long ISIs. Several studies using this approach analyzed strong thalamocortical connections and found evidence for both short-term facilitation and depression [20,21,44]. To the best of our knowledge, however, the split cross-correlogram approach has not revealed evidence of short-term plasticity in weaker synapses, such as corticocortical connections. Here we introduce two new model-based methods to characterize short-term synaptic plasticity from pre- and postsynaptic spiking. By explicitly modeling synaptic dynamics these models are able to recover a detailed description of short-term plasticity. These models reproduce the results from split cross-correlograms (Fig. 2.8), but also provide an explicit characterization of the dynamics of STP and allow reconstruction of PSP amplitudes for each presynaptic spike.

To validate our methods, we used spiking of layer 2/3 pyramidal neurons in vitro induced by injection of a current composed of PSCs from an artificial population of presynaptic neurons, whose spiking and plasticity parameters are known. Even though each presynaptic input represents only a small fraction of the total injected current, we can accurately estimate the synaptic dynamics from pre- and postsynaptic spiking. In this setting, both model-based methods, the TM-GLM and GBLM, can robustly distinguish between different types of STP, and can reconstruct PSP amplitudes for a wide range of synaptic weights for both excitatory and inhibitory connections. The TM-GLM provides a compact description of STP with four parameters related to the vesicular release and calcium dynamics in the presynaptic terminal. The GBLM provides a functional description of how the synaptic weight changes as a function of presynaptic ISIs. An advantage of the GBLM approach is that the synaptic modification rule is not constrained by the biophysics of single synapses, but has the potential to capture more complex dependences, including polysynaptic effects. One further advantage of the GBLM over the TM-GLM model is that the synaptic dynamics are assumed to be linear, which increases both the speed and robustness of the optimization process. Depending on whether a functional or a biophysical description is required, the two methods may thus both be useful tools for large-scale characterization of short-term synaptic plasticity from spiking activity.

Estimating synaptic plasticity from *in vivo* multi-electrode recordings of spiking activity will introduce several additional challenges. One challenge is simply detecting the connections between neurons. Strong monosynaptic connections are typically expressed in cross-correlograms as clear peaks (or troughs, for inhibition) with short latency and sharp onset, but weak connections or connections between neurons with low firing rates are difficult to detect in cross-correlograms. In previous work, we showed that model-based approaches can increase the sensitivity of detection

for weak connections compared to traditional cross-correlation approaches [26], and the GLM-based approaches here are likely to have similar advantages.

A second challenge is that short-term synaptic plasticity isn't the only source of variation in the observed postsynaptic responses to presynaptic spikes. Changes in the excitability of the postsynaptic neuron, stochasticity of vesicle release, and spike sorting errors can alter the statistics of the response and could potentially bias our estimates of short-term synaptic plasticity. To study how these sources of variability affect estimation of STP parameters we simulated spike trains of connected leaky integrate-and-fire model neurons, and introduced each of these confounding variables individually. We found that adding an after-hyperpolarization current (I_{AHP}) to the postsynaptic neuron impacts only the post spike-history filters in both the TM-GLM and GBLM, and does not substantially change STP estimation. Stochastic vesicle release and spike sorting errors, on the other hand, lead to biases in the estimation of short-term synaptic plasticity for our models. However, even with these additional noise sources, both the TM-GLM and GBLM are still able to reliably distinguish between connections expressing short-term facilitation and depression.

A third challenge is that correlations between the spiking of two neurons may be produced by common input rather than, or in addition to, the synaptic connection between the neurons. In our experiments with current injection in neurons in slices, inputs were generated as independent inhomogeneous Poisson processes, without the correlations that are present *in vivo*. To understand how correlated spiking can affect STP estimation, we simulated a small, feed-forward network of neurons with common input. We found that as the common input becomes stronger, the synchronization between pre- and postsynaptic spikes can interfere with the estimation of STP.

The TM-GLM and GBLM were able to estimate synaptic dynamics only when common input was weak, but failed to accurately estimate the underlying synaptic dynamics for neurons with strong common input. While our in vitro experiment and simulations allowed us to compare STP estimation under controlled conditions with known synaptic dynamics, more work may thus be needed to account for all the dependencies that occur between pre- and postsynaptic neurons *in vivo*.

Finally, a fourth challenge is that the assumptions of the TM model itself do not necessarily describe the dynamics of all interactions between the pre- and postsynaptic neurons. The TM model only aims to describe presynaptic mechanisms of STP. However, postsynaptic factors such as desensitization or saturation of postsynaptic receptors may play a role in STP at some synapses, and the synaptic weight may vary on other timescales (e.g. due to LTP/LTD). Replacing the TM model with alternative models of plasticity may be a tractable approach to address these challenges [12,14,34]. Alternatively, since the GBLM is not constrained by single-synapse biophysics, it may, in some cases, provide a more flexible first-order description of short-term dynamics, including those that are not well described by the TM model.

Rather than describing anatomical connectivity, the two model-based methods introduced here describe the plasticity of functional interactions between neurons. Many of the techniques that have been used to improve models of functional connectivity without plasticity can be used to improve the TM-GLM and the GBLM presented here. For instance, it may useful to model multiple inputs simultaneously or to include latent common input in the model [61–63]. More structured regularization techniques may allow more accurate reconstruction with smaller sets of data [64,65]. To improve models of synaptic dynamics it may be useful to consider additional

timescales [34], a higher-order expansion of the ISI dependencies, or other types of plasticity occurring on longer time scales, such as spike-timing dependent plasticity [13,66,67]. Applying these methods *in vivo* may then allow us to characterize short-term plasticity during natural behavior and in larger populations than previously possible.

Acknowledgements

We are grateful to Vladimir Ilin for help in setting up the experiments. This work was partially supported by NSF CAREER 1651396 to IS and NIH grant RO1 MH087631 to MV. AM was supported by Russian Foundation for Basic Research (grant #15-04-06286) and by the Russian Science Foundation (grant #16-15-00291, whole cell recording).

Chapter 3 (Functional connectivity with short-term dynamics explains diverse patterns of excitatory spike transmission in vivo) represents work under review:

"Functional connectivity with short-term dynamics explains diverse patterns of excitatory spike transmission in vivo" *preprint available in bioRxiv*.

In addition to the dissertation author, the following individuals also made contributions to the writing contained in this chapter:

Ian H. Stevenson[1,2]

Naixin Ren[2],

Christian Keine[3],

Carl Stoelzel[2],

Bernhard Englitz[4],

Harvey Swadlow[2]

Affiliations:

1) University of Connecticut, Department of Biomedical Engineering

2) University of Connecticut, Department of Psychological Sciences

3) Carver College of Medicine, Iowa Institute of Neuroscience, University of Iowa

4) Radboud University, Donders Institute for Brain, Cognition and Behavior, Department of Neurophysiology

Chapter 3: Functional connectivity with short-term dynamics explains diverse patterns of excitatory spike transmission *in vivo*

Short-term synaptic plasticity (STP) causes the effect of presynaptic spikes on a postsynaptic neuron to vary on timescales ranging from a few milliseconds to a few seconds. STP has been extensively studied *in vitro* by stimulating a presynaptic input with pulses of different frequencies and observing depression or facilitation in the postsynaptic potentials or currents. These studies have shown that the type and timescale of STP varies by cell type and brain region, and such differences may underlie differences in neural computations including temporal filtering of inputs, gain control, network stability, sound localization by coincidence detection, and working memory. However, since recording postsynaptic potentials (PSP) or currents (PSC) *in vivo* is challenging, short-term synaptic plasticity has not been fully characterized in awake, behaving animals. Rather than observing PSP/PSCs directly, we here deduce STP parameters from spike observations along, using a model of the translation from presynaptic to postsynaptic spiking. We model the short-term changes in the probability of a postsynaptic spike following a presynaptic spike – the synaptic efficacy. In cross-correlograms between pre- and postsynaptic spiking, monosynaptic connections show a rapid, transient change in the probability of evoking a postsynaptic spike, at a short delay after the presynaptic spike (~ <4 ms). Previous work has argued that, in depressing synapses, this probability or efficacy is larger when presynaptic spikes are preceded by long inter-spike intervals (ISIs), and in facilitating synapses efficacy is larger for short intervals. However, in practice, the observed correlation between pre- and postsynaptic spiking is a mixture of multiple underlying

phenomena. Here we develop a model-based approach for decomposing the short-term changes in the probability of a postsynaptic spike into four components: 1) short-term synaptic plasticity, 2) integration of PSPs, 3) history effects from previous postsynaptic spikes, and 4) slow common input to both pre- and postsynaptic neurons. The observed spike probability depends on each of these factors as well as the synaptic strength itself and the distribution of presynaptic spike times. We developed an extension of a typical generalized linear model (GLM) to use only pre- and postsynaptic spike observations. Our dynamical GLM allows us to characterize short-term synaptic dynamics of a wide range of synaptic behaviors *in vivo*. The estimated synaptic parameters (e.g. membrane time constant) as well as plasticity parameters (e.g. release probability, facilitation/depression time constants) could be compared with *in vitro* measurements. To validate our model, first we measure postsynaptic in three strong putative synapses response probability in a range of presynaptic inter-spike intervals and compare it with the model prediction of the response probability. Our model captures diverse patterns of spike transmission probability, disentangles them into the above factors, estimates biologically meaningful parameters of the dynamical synapse, and measures improvements in prediction of postsynaptic spiking using Receiver Operating Curves (ROC). We then demonstrate how presynaptic spiking activity beyond the most recent ISI and the postsynaptic spiking activity is affecting the spike transmission probability.

Introduction

Neural information processing is largely governed by synapses and their dynamics [2,68]. Short-term synaptic plasticity (STP) alters synaptic transmission on timescales from a few milliseconds to several seconds depending on the sequence of presynaptic spiking. Presynaptic STP arises from a mixture of two main processes: depletion of neurotransmitter-containing vesicles, which causes depression, and the elevation of residual Ca^{2+} in the presynaptic terminal, which causes facilitation by increasing vesicle release probability [69]. This can be observed in intracellular recordings where, following repetitive stimulation of the presynaptic terminal, the amplitudes of postsynaptic potentials (PSPs) or currents (PSCs) will either decrease (depression) or increase (facilitation) [69,70]. The degree of STP differs depending on the pre- and postsynaptic cell type [71] and brain region [72,73]. Functionally, STP can act as a temporal filter [74], can allow neural circuits to specialize for specific tasks [8,75], and may also underlie gain control [76], network stability [7], and long-term synaptic plasticity [77]. Here we focus on understanding how STP-induced changes in PSP/PSC amplitudes shape postsynaptic spiking. *In vivo* studies have shown that postsynaptic spiking probability, similar to the amplitude of PSP/PSCs, depends on the recent history of presynaptic spiking [20,78]. Just as PSP/PSCs show diverse patterns of depression and facilitation, the postsynaptic spiking probability also appears to have complex patterns depending on the brain region and cell-types [22]. However, postsynaptic spiking probability is modified by many additional variables besides STP at a single synaptic input. Here we aim to understand how the pattern of presynaptic spiking activity and short-term synaptic plasticity shape postsynaptic spiking probability.

To do so, we use simultaneously recorded pre- and postsynaptic spiking activity to detect and study putative monosynaptic connections. When an excitatory, monosynaptic connection is present,

cross-correlations between the spiking of a pre- and postsynaptic neuron often show a rapid, transient increase in postsynaptic spikes following the presynaptic spike. This occurs at an interval reflecting the presynaptic axonal conduction time plus the synaptic delay (usually < 4 ms) [19,79]. However, this cross-correlation is not static. Previous studies have found that the cross-correlation often differs for presynaptic spikes that are part of a burst compared to isolated spikes [20]. Spike transmission probability appears to depend on the timing of previous presynaptic spikes, and one factor influencing spike probability may be STP [80,81]. For example, depressing synapses would have more reliable synaptic transmission in response to isolated presynaptic spikes following long inter-spike intervals (ISIs) compared to shorter intervals (in bursts) [20,22]. On the other hand, facilitating synapses would show a stronger response to presynaptic spikes following shorter ISIs (bursts) compared to the presynaptic spikes following longer ISIs (isolated) [44]. By looking at the corresponding cross-correlograms from a subset of presynaptic spikes with specific ISIs, previous studies have found highly diverse, non-monotonic spike transmission patterns for different synapses [22]. This diversity in patterns of spike transmission probability, however, is not solely attributable to STP. When two presynaptic spikes occur in close succession, the membrane time-constant may cause postsynaptic potentials (PSP) to sum and increase the spike probability even if the individual PSPs were sub-threshold [82]. Moreover, the history of postsynaptic spiking also affects spike probability such that even if the PSP is strong, it may not trigger a spike if it falls within the refractory period or during an after-hyperpolarization (I_{AHP}) current [83]. Finally, slow fluctuations in the overall excitability of the postsynaptic neuron, due to neuro-modulation, for instance, could also change synaptic transmission probability [84]. In different synapses the degree that each factor contributes varies and leads to diverse patterns of postsynaptic spike transmission probability.

The overall correlation structure in spiking data can often be estimated by generalized linear models (GLMs) [24,49]. However, previous models have treated these interactions as static, and, thus, cannot capture dynamic changes in spike transmission probability. Here we extend these GLMs to describe dynamic interactions between neurons and account for the diversity of spike transmission patterns [24,85,86]. For each individual presynaptic spike, our model aims to predict postsynaptic spikes accounting for the postsynaptic neuron's baseline firing rate, slow fluctuations of the postsynaptic firing rate, the effect of postsynaptic spiking history, and a coupling term affected by synaptic summation and short-term synaptic plasticity. The conditional intensity of our model provides estimates of postsynaptic spiking probability following every single presynaptic event based on the previously observed sequence of pre- and postsynaptic spiking. The split cross-correlogram only describes the average response conditioned on the ISI preceding the most recent presynaptic spike. By using a model-based approach we can incorporate the full sequence of presynaptic spikes beyond the most recent one, explicitly account for factors such as postsynaptic history, and link the observed patterns of spike transmission to the underlying dynamics of vesicle depletion and release probability.

To evaluate the model, we first examined its ability to capture the observed patterns of spike transmission probability for three well-studied, strong putative synapses using pre- and postsynaptic spike observations from: 1) a pair of neurons in the mouse thalamus, 2) an auditory nerve projection onto a spherical bushy cell (ANF-SBC) in the gerbil, and 3) a neuron in ventrobasal (VB) thalamus of the rabbit projecting to a putative fast-spike inhibitory neuron in primary somatosensory (S1) barrel cortex (VB – Barrel). Short-term synaptic dynamics of this latter system have been extensively characterized *in vivo* [20,87,88]. Similarly, ANF-SBC synapses have been extensively studied in previous experiments and are well-characterized *in vitro*

[89–91]. Using the auditory brainstem connection, we explore how synaptic transmission probability changes depending on the stimulus. Our result suggests that a simplified model, without considering short-term plasticity, is insufficient to explain how patterns of spike transmission change as the pattern of presynaptic input changes. Finally, we apply our model to spiking data from a large-scale, multi-electrode array recorded from multiple areas in an awake mouse. Here we investigate the STP dynamics in putative synapses from excitatory neurons onto two putative inhibitory subtypes (e.g. FS: fast-spiking, RS: regular-spiking). We find that these two types of connections have distinct patterns of spike transmission, where excitatory-FS connections appear to be slightly more depressing than excitatory-RS connections. Together, these results illustrate the diversity of spike transmission patterns *in vivo* and present one potential approach to studying short-term synaptic dynamics in behaving animals.

Most previous approaches to describing interactions between neurons using spiking activity have focused on static functional connectivity. These models improve both encoding and decoding accuracy and have been shown to capture physiological network structure in some cases [25]. Here we model dynamic functional connectivity where the effect of each presynaptic spike on the probability of postsynaptic spiking depends on the previously observed sequence of presynaptic spiking. This augmented GLM can be directly compared with the observed spike transmission probability and also allows us to disentangle the contributions of short-term synaptic plasticity, synaptic summation, presynaptic firing rate fluctuations, and spike history. Moreover, we find that modeling dynamic functional connections allows us to better predict postsynaptic responses compared to the static models. Since modeling static functional connectivity can improve decoding [24,49,92], modeling dynamic functional connectivity may improve decoding further as well. As multi-electrode arrays improve, and the number of simultaneously recorded neurons increases,

models of dynamic functional connectivity may provide insight into not just network structure, but also the extent of depression or facilitation in these networks, as well as differences in network dynamics across multiple brain areas and under different behavioral conditions. Here, our findings suggest that, at least in some cases, *in vivo* spike transmission dynamics differ substantially for different stimuli and different cell-types.

Methods

Neural Data

To illustrate how synaptic dynamics can be estimated from spikes, we first examined a set of three strong putative or identified synapses with diverse spike transmission probability patterns: (i) a dual-electrode recording of a thalamocortical projection in the barrel system, (ii) an *in vivo* loose-patch (juxtacellular) recording at the calyceal endbulb of Held synapse in the auditory brainstem, and (iii) a recording from a pair of neurons in the thalamus detected from a larger multi-electrode array (MEA) recording. Next, we applied our model more generally to analyze a large sample of putative synaptic connections recorded from the MEA dataset. The data from these three identified strong synapses and the MEA data have been collected from different species, regions, cell-types, under different stimulation and show a diverse pattern of postsynaptic spiking probability. In all cases we deduce short-term synaptic dynamics on the basis of only pre- and postsynaptic spike observations.

For the first putative synapse, we use *in vivo* data from simultaneous extracellular recordings in ventrobasal (VB) thalamic barreloids and topographically aligned, somatosensory cortical barrel columns (VB-Barrel) in awake, unanesthetized, adult rabbits. Detailed surgical and physiological methods have been described previously [93]. Spike-triggered averages of the cortical spikes following spiking of the VB neuron was used to identify connected S1 neurons. Based on the presence of high frequency discharge (3+ spikes, > 600 Hz) following electrical stimulation of the thalamus, and narrow spike waveforms, the S1 neuron in this recording was identified as a putative inhibitory neuron [94]. These recordings identified several putative thalamocortical projections. The putative synapse that we model here is particularly clear, with 68,345 pre- and 128,096

postsynaptic spikes recorded over the course of 92 minutes of spontaneous activity and has been previously studied in [20,21].

For the second synapse, we examined *in vivo* loose-patch recordings at the Endbulb of Held in young adult gerbils. Detailed surgical and physiological methods have been previously described [95]. Briefly, the glass electrode was positioned in the anterior portion of the ventral cochlear nucleus (AVCN) and single-units were recorded during varying acoustic stimulation. Single units were classified when recording a positive action potential amplitude of at least 2 mV and showing the characteristic complex waveform identifying them as large spherical bushy cells (SBC) of the rostral AVCN. This recording included a mixture of juxtacellular waveforms: an isolated excitatory PSP (EPSP) or an EPSP followed by a postsynaptic action potential. For both cases the timing of EPSPs and spikes and rising slope of the EPSPs were extracted. The timing and slope of the EPSPs were identified using a slope threshold for the rising part of EPSPs as previously described [96]. We then modeled spike transmission probability patterns for two recordings: (i) during randomized pure tone acoustic stimulation and (ii) during multiple stimuli, i.e. randomized frequency-level pure tone stimulation interspaced with spontaneous activity, natural sounds, and also during spontaneous activity. Using this second dataset, we characterized how variable presynaptic spike patterns evoked by different stimuli affected the patterns of spike transmission at the same synapse.

We also use MEA spiking data to study the factors shaping spike transmission probability patterns in a large-scale recording with multiple cell-types. Here we use a previously collected, publicly available recording from the Cortex Lab at UCL [97,98] with data from two Neuropixels electrode arrays recorded simultaneously, each with 960 sites (384 active) with lengths of 10-mm and spacing of 70 × 20-μm (http://data.cortexlab.net/dualPhase3/). The two electrode arrays span

multiple brain areas and ~90 min of data was collected in an awake, head-fixed mouse on a rotating rubber wheel during visual stimulus presentations. Spikes were automatically detected and sorted using Kilosort [99] on the broadband (0.3–10 kHz) signal and then manually curated. If two clusters of spikes had similar waveforms, cross-correlogram features, and spike amplitudes, they were merged into a single cluster and assigned to a single neuron. In total, 831 well-isolated single neurons where identified from the two probes in several different brain areas: visual cortex (n=74), hippocampus (n=64), thalamus (n=244), motor cortex (n=243), and striatum (n=200). Due to the large number of simultaneously recorded neurons in this dataset, there are many potential synapses (~831^2).

Synapse Detection:

To identify putative monosynaptic connections between well-isolated single neurons, we looked for specific patterns in the cross-correlograms [100]. If two neurons are monosynaptically connected, the probability of postsynaptic spiking increases/decreases rapidly following a presynaptic spike. In spiking data, this rapid, transient change can be seen in cross-correlograms as an asymmetric bump/dip in the number of postsynaptic spikes following presynaptic spikes [79]. For each connection we calculated the cross-correlogram in a 5 ms window before and after presynaptic spikes with bin-size of 0.1 ms. To avoid aliasing in the cross-correlograms, we added a small, random shift to each postsynaptic spike drawn uniformly between $-\Delta t/2$ and $\Delta t/2$ where Δt is the spike time resolution (0.01 ms in most cases). Here we used a model-based approach using the cross-correlograms to decide whether two synapses are monosynaptically connected. To fit the cross-correlogram we used a baseline rate μ, a linear combination of B-spline bases $\boldsymbol{B}(t)$, and a weighted alpha function to model the synapse, $w\,\boldsymbol{\alpha}(t)$, all passed through an output nonlinearity; $\lambda(t) = \exp\big(\mu + \boldsymbol{r}\boldsymbol{B}(t) + w\,\boldsymbol{\alpha}(t)\big)$. The alpha function, $\boldsymbol{\alpha}(t) = (t - t_d)/$

$\tau_\alpha \exp(1-(t-t_d)/\tau_\alpha)$, describes the shape of the synaptic potential where t_d is the synaptic delay and τ_α is the synaptic time-constant [82]. For individual connections, we estimate these parameters by maximizing the penalized Poisson log-likelihood $l(\mu, \boldsymbol{r}, w, t_d, \tau_\alpha) = \Sigma y_i log\lambda_i - \Sigma\lambda_i + \epsilon\|\boldsymbol{r}\|_2$ where y_i is the number of postsynaptic spikes observed in the i-th bin of the correlogram and $\|\boldsymbol{r}\|_2$ regularizes the model to penalize B-spline bases for capturing sharp increases in the cross-correlogram. ϵ is a regularization hyper-parameter which we set to 1 based on manual search. Due to the parameterization of $\boldsymbol{\alpha}(t)$, the log-likelihood is not concave. However, since the gradient of the log-likelihood can be calculated analytically, we efficiently optimize the likelihood using LBFGS. During the optimization, the delay and time-constant parameters are log-transformed, allowing us to use unconstrained optimization, even though they are strictly positive. We used random restarts to avoid local maxima. To identify putative monosynaptic connections in the large-scale multi-electrode array data, we compared this model with a smooth model with slow changes in cross-correlogram and without the synapse, $\lambda_0(t) = \exp(\mu' + \boldsymbol{r'B}(t))$, using the log-likelihood ratio (LLR) test between our full model with synapse and the nested smooth model. Since low values of the likelihood ratio mean that the observed result was better explained with full model as compared to the smooth model, we then visually screened pair-wise connections with lowest ratios (LLR < -6) compared to the null model to find putative synapses. Out of ~831^2 possible connections in this dataset we find ~200 putative synapses (0.03%). We handpicked a strong putative synapse between two thalamic neurons to study its efficacy pattern in detail alongside the VB-Barrel and ANF-SBC synapses.

In addition to this single strong synapse, we also categorize putative pre- and postsynaptic cell types for the connections detected in the MEA dataset. For this purpose, we assessed single units based on their cross-correlograms, firing rates, and spike waveforms. We categorized units as

excitatory or inhibitory if, in accordance with Dale's law [101], all outgoing cross-correlograms showed transient, short-latency (<4ms) increase/decrease in spiking probability [92]. We then looked into identified inhibitory neurons and categorized them into to putative fast-spiking (FS) and regular-spiking (RS) inhibitory neurons. Using these putative Excitatory-FS and Excitatory-RS synapses, we then examine how the spike transmission patterns differ for these two subtypes of inhibitory neurons.

Extending a Generalized Linear Model to Account for Short-term Plasticity (TM-GLM)

Short-term synaptic plasticity causes the amplitude of postsynaptic potentials (PSP) to vary over time depending on the dynamics of synaptic resources and utilization and can be modeled using the pattern of presynaptic spiking [11,32]. However, changes in the overall postsynaptic spiking probability cannot be uniquely attributed to changes in amplitudes of postsynaptic potentials. To accurately describe the dynamics of spike transmission, we also need to account for the membrane potential summation, the excitability of the postsynaptic neuron (e.g. slow changes in the presynaptic firing rate) and the dynamics of postsynaptic spiking (e.g. refractory period, after hyperpolarization current). We developed an extension of a generalized linear model, which we call a TM-GLM to describe each of these effects. Concretely, the probability of a postsynaptic spike shortly after each presynaptic spike accounts for the full sequence of previous presynaptic spiking and the recent history of postsynaptic spiking. We define the conditional intensity of the postsynaptic neuron after the i-th presynaptic spike, $t_s^{(i)}$, so that the probability of observing a postsynaptic spike in the j-th time bin after the i-th presynaptic spike is given as:

$$\lambda_i(t_j) = \sigma\left(\beta_0 + X_c\left(t_s^{(i)}\right)\beta_c + \sum_{t_r^{(l)}<t_s^{(i)}} X_h\left(t_s^{(i)} - t_r^{(l)}\right)\beta_h \; + \; A_s \, w_i \, \alpha(t_j)\right)$$

where $t_r^{(l)}$ are the postsynaptic spike times preceding $t_s^{(i)}$. For each presynaptic spike, our model decomposes the firing rate of the postsynaptic neuron into four effects: a baseline firing rate, β_0, slow fluctuations in presynaptic firing rate $X_c\beta_c$, history effects from the recent postsynaptic spikes (prior to $t_s^{(i)}$), $X_h\beta_h$, and a time-varying coupling effect from the presynaptic input, $A_s w\ \alpha(t)$ (Fig. 3.1).

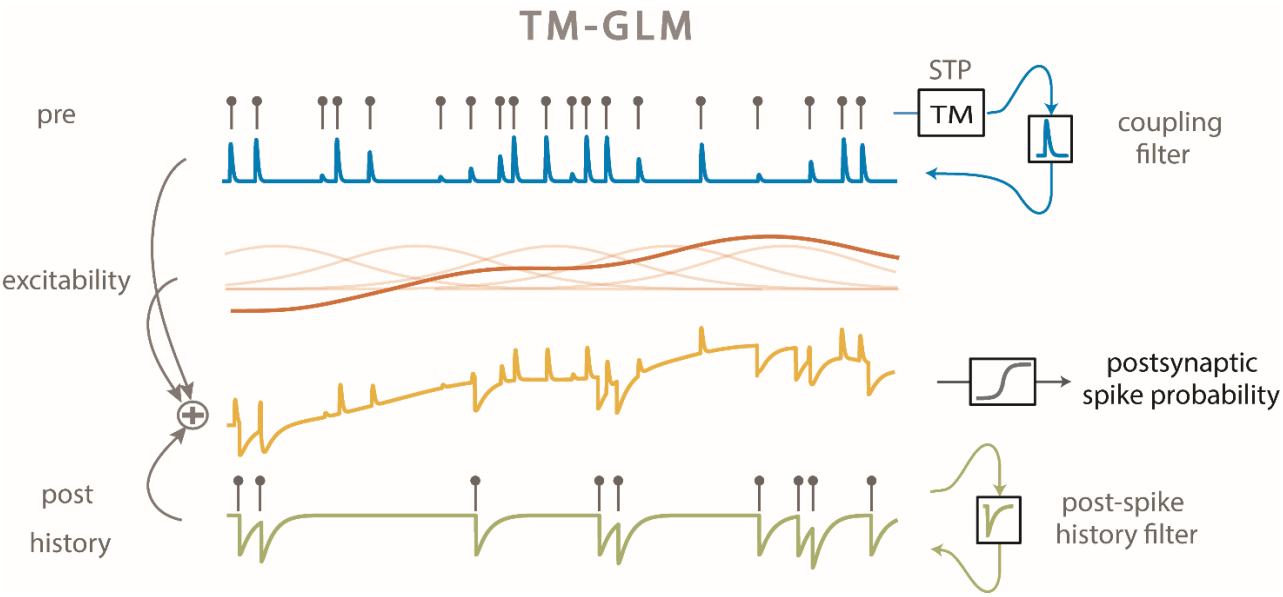

Fig. 3.1: TM-GLM. Postsynaptic spiking probability before passing the spiking nonlinearity (yellow) changes as a linear combination of presynaptic coupling term with STP dynamics (blue), postsynaptic spiking history (green), the postsynaptic excitability (red). Transparent red curves show the bases of slow changes in postsynaptic probability at presynaptic spike times (X_c).

Here we model slow fluctuations in the postsynaptic rate $X_c\beta_c$ with a linear combination of B-splines with equally spaced knots every 50 seconds of recording time. In the history term, splines (X_h) span a period of 10 ms prior to each presynaptic spike with 4 logarithmically-spaced knots. By scaling $\boldsymbol{\alpha}(t_j)$ with a multiplicative factor, w_i, the strength of a synapse can vary over time and, in this case, depends on the detailed sequence of presynaptic spiking and their corresponding inter-

spike intervals. A_s is the magnitude of the synaptic strength. In this case we use a model for short-term synaptic plasticity that allows both depression (where the w_i decreases for shorter presynaptic ISIs) and facilitation (where the w_i increases for shorter presynaptic ISIs), and incorporates membrane summation. To model these effects, w_i is determined by a nonlinear dynamical system based on the Tsodyks and Markram (TM) model [11,102] where: $w_i = w_{i-1} \exp\left(-\frac{t_s^{(i)} - t_s^{(i-1)}}{\tau_s}\right)\pi_i + R_i u_i$, where τ_s is the membrane time-constant and the first term of the equation describes how postsynaptic membrane potential summation increases the probability of postsynaptic spiking. This membrane summation will be ignored if there is a postsynaptic spike: $\pi_i = \left\{0 \text{ if } t_s^{(i-1)} < t_r^{(i-1)} < t_s^{(i)}; 1 \text{ otherwise}\right\}$. In the second term of this equation, R represents the dynamics of resources and u describes their utilization.

$$R_i = 1 - [1 - R_{i-1}(1 - u_{i-1})] \exp\left(-\frac{t_s^{(i)} - t_s^{(i-1)}}{\tau_d}\right)$$

$$u_i = U + [\, u_{i-1} + f(1 - u_{i-1}) + U] \exp\left(-\frac{t_s^{(i)} - t_s^{(i-1)}}{\tau_f}\right)$$

where τ_d and τ_f are the depression and facilitation time-constants. U is the release probability and f is the magnitude of facilitation. To make the estimation more tractable, we approximate the full optimization problem and estimate synaptic delay, t_d, and time-constant, τ_α, by fitting $\alpha(t)$ using the full cross-correlogram, as above. We fix these parameters for the rest of the optimization process. We then maximize a penalized, Bernoulli log-likelihood $l(\theta) = \Sigma\Sigma\left[y_{ij}\lambda_i(t_j) - \left(1 - y_{ij}\right)\left(1 - \lambda_i(t_j)\right)\right] + \gamma\left\|\theta'_{stp}\right\|_2$ where $\gamma = 1$ is the regularization hyperparameter to estimate the parameters: $\theta = \left\{\beta_0, \beta_{c=1:C}, \beta_{h=1:H}, A_s, \theta_{stp}\right\}$, $\theta_{stp} = \{\, \tau_d, \tau_f, U, f, \tau_s\}$.

As with previous applications of GLMs, we assume that bins are conditionally independent given the covariates, but unlike many other GLMs, here we only calculate the log-likelihood during short intervals (5ms) after presynaptic spikes. With y_{ij} being a binary value representing the presence of a postsynaptic spike in the j-th time bin after the i-th presynaptic spike. We again used a logarithmic transformation for the time-constants to avoid negative values and logit transformation for U and f to bound their values in the interval $[0,1]$; $\theta'_{stp} = \{\log(\tau_d), \log(\tau_f), \mathrm{logit}(U), \mathrm{logit}(f), \log(\tau_s)\}$. By modeling STP this model is no longer a strict GLM, and the log-likelihood may have local maxima. Here we use random restarts to avoid local maxima in our optimization process. The parameters of each restart $\{\beta_0, \beta_{c=1:C}, \beta_{h=1:H}, A_s\}$ are initialized by adding noise ($\sim N(0,1)$) to the corresponding parameters in a standard GLM. We initialize the plasticity parameters with $\tau'^{(0)}_d \sim N(-1,5)$, $\tau'^{(0)}_f \sim N(-1,5)$, $U'^{(0)} \sim N(0,5)$, $f'^{(0)} \sim N(0,5)$, $\tau'^{(0)}_s \sim N(-3,5)$. We then use an LBFGS algorithm to optimize the log-likelihood where we calculate all derivatives analytically except for derivatives of θ_{stp} which we calculate numerically. To estimate the uncertainty of the parameters, we bootstrap the data from each of the strong synapses by chunking the whole recording time into samples of 50 seconds then resampling the chunks to generate a new spike train with the same original length.

Calculating spike transmission probability

To demonstrate how the probability of postsynaptic spiking changes according to the corresponding presynaptic inter-spike intervals, we estimated spike transmission probabilities from the cross-correlograms directly instead of using a model. To calculate this probability, we focused on a transmission interval after the presynaptic spike where the conditional intensity (when corrected for the baseline rate) goes above 10% of the maximum of $\boldsymbol{\alpha}(t)$. We split the presynaptic

inter-spike interval distribution into log-spaced intervals, and, for each interval, we calculate the ratio between numbers of postsynaptic spikes in the transmission interval to the number of presynaptic spikes. Unlike previous studies [20,93] we do not correct this probability for the baseline postsynaptic rate. The uncorrected probability allows us to more directly compare the model predictions to the empirical spike transmission probabilities. Since our model gives an estimate of the postsynaptic probability after each individual presynaptic spike, we can average over the same transmission interval. However, we know if there is a postsynaptic spike in the transmission interval, probability of a postsynaptic spike goes to ~0 for all consecutive bins due to the post-spike dynamics (e.g. refractory period). Therefore, we measure the predicted probability of a postsynaptic spike in a 5ms window after i-th presynaptic spike from binned $\lambda_i(t_j)$ as follows: $z_i = \sum_{j=1}^{L} \lambda_i(t_j) \prod_{m=1}^{j-1} (1 - \lambda_i(t_m))$. Here we assume conditional independence of the j-th bin after a presynaptic spike, but we enforce a refractory period for all bins after a postsynaptic spike in our generative model. Here L is the first bin that y_{ij} is nonzero. z_i represents the probability of postsynaptic spiking after each presynaptic spike and we fit a smooth curve over the distribution of z_i's and their corresponding inter-spike intervals to compare with the empirical spike probability patterns.

Modeling the effect of local patterns of pre- and postsynaptic spiking

The observed and modeled spike transmission patterns, as calculated above, reflect the expected postsynaptic spike probability given a specific presynaptic ISI. However, since the presynaptic ISIs are not independent and there are serial correlations in ISIs, the detailed sequence of the pre- and postsynaptic spiking likely affects the shapes of these curves. To quantify the effects of serial ISI correlations on the model of spike transmission probability we demonstrate how local patterns

of presynaptic spiking modifies spike transmission patterns in the data and the model. For each of the three strong identified synapses we measure postsynaptic spiking probability in response to presynaptic spike triplets. Due to the limited number of spikes in our data, we divide the presynaptic ISI distribution into few log-spaced intervals and measure the postsynaptic spiking probability for triplets with the two ISIs that fall in those intervals. Similarly, we measure the predicted postsynaptic probability in response to the presynaptic triplets. After measuring postsynaptic responses to presynaptic spike triplets in the data and the model, we simulate the contribution of STP in shaping the transmission pattern in response to these triplets. To factor out contributions of the postsynaptic history and slow changes in presynaptic firing rate, we fix the corresponding values in the model to their average values within the model. In these simulations, we also fix the initial values of the STP dynamics in the TM model for the first spike of the triplets to the average R and u within the model. This approach enables us to illustrate how short-term synaptic plasticity in triplets of presynaptic spikes changes spike transmission probability and how serial correlations in presynaptic spiking affect spike transmission probability.

The postsynaptic spike history and the serial correlations between the pre- and postsynaptic spiking also modify spike transmission probability patterns. To investigate history effects in the local pattern of pre- and postsynaptic spikes, we measured the postsynaptic spiking probability in response to two presynaptic spikes and a postsynaptic spike preceding the most recent presynaptic spike. Due to the limited number of spikes and sparseness of the split cross-correlograms, we again divided the presynaptic and postsynaptic ISI distributions into a few log-spaced intervals. We then measure the spike transmission probability for a group of presynaptic spikes that their preceding presynaptic ISIs and postsynaptic spike ISIs fall into different combinations of pre- and postsynaptic log-spaced intervals. After measuring postsynaptic responses to any possible

combination of the two most recent presynaptic spikes and their postsynaptic spikes in the data and the model, we simulate the contribution of the history and STP together in shaping the transmission. In our simulation the excitability was set to the model estimates. To measure the effects of postsynaptic spiking history, for each postsynaptic ISI, we fix the history contribution to estimated post-spike history filter value at that postsynaptic ISI. We then use the predicted STP parameters from the data to simulate the STP contribution in response to paired pulses of presynaptic ISIs where we again fix the initial values of the TM model for the first presynaptic spike to the average R and u within the model. This approach enables us to illustrate how short-term synaptic plasticity in local patterns of two presynaptic spikes and a postsynaptic spike changes spike transmission probability and quantifies how serial correlations between pre- and postsynaptic spiking affect spike transmission probability.

Evaluating prediction accuracy

In addition to evaluating the estimated parameters and comparing the model to empirical spike transmission probabilities, we also assess how accurately the model can predict postsynaptic spiking. Not only can we predict the probability of a spike given specific presynaptic ISIs, but we can also predict whether there will be a postsynaptic spike following each individual presynaptic spike. To quantify how well the predicted postsynaptic spike probability, z_i, predicts the postsynaptic spiking activity, we use Receiver Operating Characteristic (ROC) curves. To compute the ROC curve, we first create a threshold version of z_i which operates as our prediction: $\{(\hat{r}_i = 1)$ if $(z_i > \mathrm{thr})$; 0 otherwise$\}$. Changing the threshold from 0 to 1 traces out a relationship between the true positive rate (TPR) and false positive rate (FPR). The area under the ROC curve (AUC) reflects the performance of each model, where a perfect classifier has AUC=1 and a random classifier has AUC=0.5. Effectively, the AUC is the probability of a randomly chosen spike having

a higher model probability than a randomly chosen non-spike [103]. Here we calculate the AUC for short intervals (~5ms) after presynaptic spikes and check whether we detect a postsynaptic spike in the transmission interval where $\alpha(t)$ is above 10% of its maximum. Here we compare the AUC for the static model of connectivity without short-term synaptic plasticity with our dynamical model.

A simplified rate model to simulate effects of synaptic summation and post-spike history

Our TM-GLM's prediction of the spike transmission pattern is data-driven and depends on the full history of pre- and postsynaptic spiking. To better understand and illustrate how STP, synaptic summation, and post-spike history interact to create the observed patterns of spike transmission, we simulated postsynaptic responses in a simplified voltage model. Namely, we consider PSP summation in response to a pattern of two presynaptic spikes. We assume that the synapse is initially fully recovered, and the PSC amplitudes are determined by the 4-paramter TM model with $U = 0.7$, $\tau_d = 1.7$, $\tau_f = 0.02$, f =0.05 for the depressing synapse and U =0.1, $\tau_d = 0.02$, $\tau_f = 1$, f =0.11 for the facilitating synapse [86]. We then convolve the PSCs (delta function kernel) with a PSP kernel, $\exp(-t/\tau_v) - \exp(-t/\tau_r)$, with $\tau_v = .01$ and τ_r=.001 ms to describe synaptic summation. We assume that the instantaneous postsynaptic spike probability is simply a nonlinear function of the distance to a threshold voltage $\sigma\big(5(V(t) - V_{th})\big)$ where $\sigma(x) = 1/(1 + e^{-x})$ and $V_{th} = .5, .75$, and 1 correspond to strong, moderate, and weak inputs respectively. The spike transmission probability sums this instantaneous probability over a window of 20ms after each presynaptic spike. Finally, we adjust the spike transmission probability for the second PSP to account for potential post-spike history effects. Namely, we assume that the adjusted spike transmission probability for the second spike is $p_2^* = (1 - p_1)p_2 + p_1 p_2 f_{ahp}$ where p_1 is the

transmission probability for the first spike, p_2 is the unadjusted probability for the second spike, and f_{ahp} is the effect of the after-hyperpolarization. Here we use $f_{ahp}(\Delta t) = (\sigma\big(150(\Delta t - 0.02)\big) - c)/d$ where Δt is the presynaptic ISI, and c and d are constants ensuring that $f_{ahp}(0) = 0$ and $f_{ahp}(\infty) = 1$. Although this simulation is highly simplified, it demonstrates how the observed spike transmission pattern depends, not just on the type and timescale of STP, but on the interaction between STP, synaptic summation, after-hyperpolarization effects, and the spike nonlinearity.

Results

Short-term synaptic plasticity directly affects synaptic information processing by altering the amplitude of presynaptic currents [68], but in most neural systems it remains unclear how these presynaptic effects translate to altered postsynaptic spike probability. Postsynaptic spiking is affected by many factors including short-term plasticity, postsynaptic spike history, summation of PSPs, and slow fluctuations in presynaptic firing rate. Here we developed a statistical model that includes each of these factors and allows their effects to be estimated solely using pre- and postsynaptic spiking activity. In this approach, working with spikes rather than PSC/PSPs enables us to understand the short-term changes in synaptic transmission probability *in vivo* where large-scale intracellular recordings have not been achieved.

Spike transmission probability varies strongly as a function of presynaptic ISIs

Here we define spike transmission probability as the probability of postsynaptic spiking in a window shortly after each presynaptic spike. One conventional approach to study spike transmission and changes in transmission probability are cross-correlograms. Cross-correlograms of excitatory monosynaptic connections show a rapid, transient increase in the postsynaptic spiking probability shortly (usually < 4ms, although this depends on the presynaptic axonal conduction delay) after the presynaptic spike [19]. The timing and shape of the cross-correlogram depends on the synaptic delay, the strength of the connection and varies between synapses. However, in the overall cross-correlogram since all presynaptic ISIs are averaged, the dependence of spike transmission probability on the presynaptic ISIs remains hidden (Fig. 3.2A). To determine the effect of presynaptic ISI on spike transmission probability we can calculate the cross-correlogram for a subset of presynaptic spikes with a specific ISI. and previous studies showed that transmission probabilities can vary for different ISIs within the same synapse [20,22]. Moreover, the short-term

dynamics of spike transmission probability can differ for different synapses as a function of presynaptic ISIs. To illustrate this diversity, we examined three strong synapses from three distinct neural systems: (i) a pair of neurons in thalamus in a mouse, (ii) a projection from ventrobasal to somatosensory barrel cortex (VB-Barrel) in a rabbit, and (iii) the auditory nerve fiber to spherical bushy cell projection in a gerbil (ANF-SBC). Although the presynaptic neurons have diverse ISI distributions (Fig. 3.2B), splitting the spikes into ISI quantiles and calculating the correlogram for each quantile, demonstrates how postsynaptic responses differ following short and long presynaptic ISIs. For the pair of neurons in thalamus, spike transmission has a higher probability at short and long intervals and a lower probability for mid-range ISIs. For VB-Barrel transmission probability is higher for longer ISIs, while for ANF-SBC the highest transmission probability occurs at intermediate intervals (Fig. 3.2C).

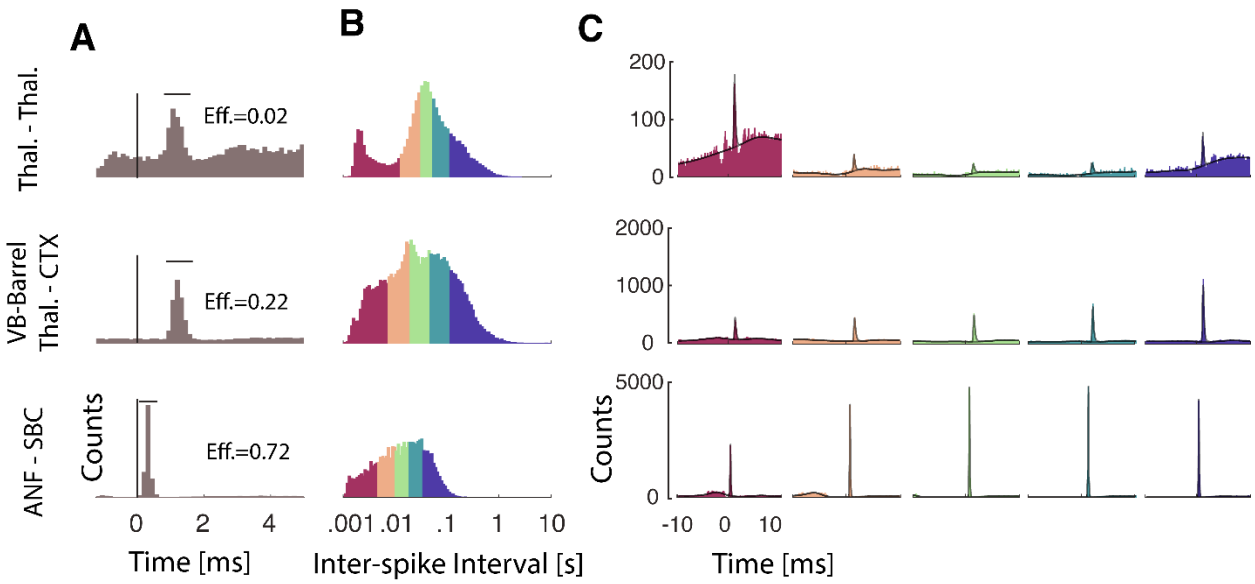

Fig. 3.2: Spike transmission probability depends on the presynaptic ISI and differs between synapses. A) *Cross-correlograms between pre- and postsynaptic spiking show an increase in the postsynaptic spike count (or probability) after a short latency, indicative of a monosynaptic connection. Efficacy (Eff.) for each synapse denotes the ratio between numbers of postsynaptic spikes in the synaptic peak (horizontal bars) corrected for the baseline number of expected postsynaptic spikes to number of presynaptic spikes.* ***B)*** *Inter-spike interval distributions (log-scale) for the presynaptic neuron for three different synapses. The distributions are color-coded into 5 quantiles with equal number of presynaptic spikes.* ***C)*** *For each ISI quantile, we calculate a separate cross-correlogram. Colors correspond to (B) going from shorter presynaptic ISIs (left) to longer ISIs (right). Note that both the baseline firing rate and the synaptic peak for each connection change as a function of presynaptic ISI. Solid lines overlaying the cross-correlograms illustrate model fits used to estimate the synaptic effect and the smooth baseline.*

The shape of spike transmission patterns depends on multiple factors

In synapses exhibiting short-term synaptic plasticity the postsynaptic response after each presynaptic spike changes according to the recent history of presynaptic spiking [11,86]. However, besides synaptic dynamics there are additional factors that alter spike timing. At short presynaptic ISIs, membrane potential summation can lead to larger PSPs and increased spike probability, even in absence of short-term synaptic plasticity [82]. The spiking nonlinearity and the history of postsynaptic spiking can also alter how a given pattern of presynaptic input is transformed into postsynaptic spiking [24,104].

To illustrate how STP, synaptic summation, and postsynaptic history interact to create the observed spike transmission pattern we simulated from a simplified rate model with linear voltage summation, short-term plasticity, a soft spiking nonlinearity, and an after-hyperpolarization (Fig. 3.3). Similar to experimental data, the spike transmission probability in this simplified model depends on the presynaptic ISI as well as the type of STP. For depressing synapses, the spike transmission probability increases for longer presynaptic ISIs while for facilitating synapses it increases for mid-range ISIs [11,86]. Independent of STP type, PSPs sum at short ISIs (Fig. 3.3A). However, the exact shape of transmission probabilities also depends on the strength of the synapse and, possibly, the history of postsynaptic spiking. An after-hyperpolarization current following each postsynaptic spike, for instance, can briefly decreases the probability of subsequent spikes. In our simulation, we find that "spike interference" from previous postsynaptic activity can counteract membrane potential summation (Fig. 3.3B). This type of postsynaptic spike interference generally decreases the spike probability for shorter presynaptic ISIs, but the magnitude of this decrease depends on the synaptic strength and type of STP (Fig. 3.3C). These simulations illustrate how the pattern of spike transmission probability results from the complex

interaction between the membrane potential, the spike nonlinearity, the post-spike history, and short-term synaptic plasticity.

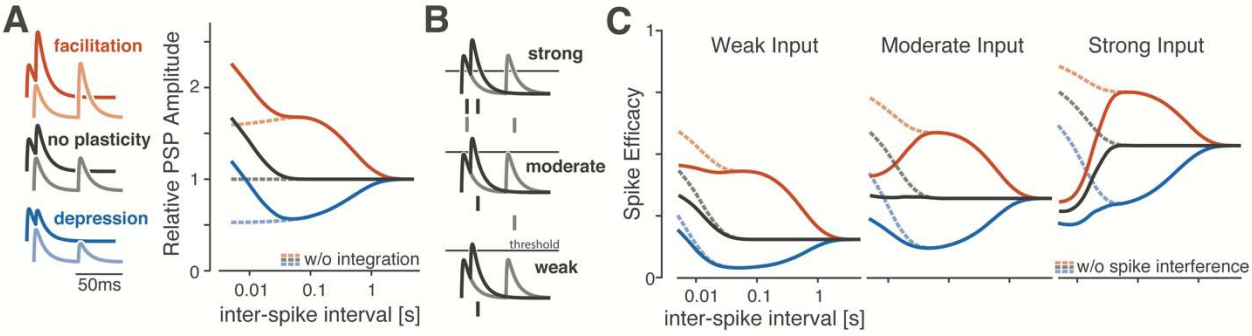

Fig. 3.3: A simulation of a simplified rate model shows how spike transmission probability depends on multiple factors. A: *For different types of short-term synaptic plasticity, postsynaptic summation increases the amplitudes of the postsynaptic potentials (PSP) at shorter ISIs. Lines denote the membrane potential of a postsynaptic neuron in a simplified model as it responds to short (dark traces) and long (light) paired presynaptic pulses. Relative amplitudes of excitatory postsynaptic potentials increase under the simplified model depending on the type of STP (right panel).* ***B:*** *Spike generation changes with synaptic strength. In this paired-pulse stimulation paradigm, stronger synapses are more likely to generate a spike following the first presynaptic impulse which can then decrease the spiking probability following the second impulse if there are post-spike history effects. As in (B) traces denote postsynaptic membrane potential responses to short (dark) and long (light) presynaptic ISIs. Dashes denote example postsynaptic spiking, with "spike interference" occurring for strong synapses and short ISIs.* ***C:*** *The pattern of spike transmission probability under the simplified model changes depending on the type of STP, the coupling strength, and presence of post-spike interference. Dashed lines show transmission probability without interference from previous postsynaptic spikes, while solid lines show how post-spike history effects can decrease the spike transmission probability.*

Spike transmission patterns are diverse across regions and species

The combination of these factors could be one explanation for to the diversity of spike transmission patterns in experimental data. To account for STP, postsynaptic history effects, and slow changes of firing rate we extend a previously developed GLM framework for static functional connections [24] to include short-term dynamics. In the previous, static GLM the probability of postsynaptic spiking is modeled as a linear combination of a baseline firing rate parameter, a post-spike history filter to capture the postsynaptic spike dynamics, such as refractoriness and burstiness, and a coupling filter describing the fixed influence of presynaptic spikes. The sum of these effects is then passed through a spiking non-linearity [24]. In our extended model we added a linear term that allows changes in excitability of the postsynaptic neuron as a function of the presynaptic firing rate (timescale >1min) and allow the coupling term to change for each presynaptic spike according to the Tsodyks and Markram (TM) model of STP [11]. We fit the parameters of this TM-GLM using only the pre- and postsynaptic spike observations and obtain parameters for each effect using approximate maximum likelihood estimation (see Methods). This provides estimates of the history and coupling filters, as in a static GLM, as well as additional parameters for the dynamical synapse (TM model) including facilitation, depression, membrane time constants, and release probability. Given these parameters, the model estimates the postsynaptic spiking probabilities following each observed presynaptic spike and predicts spike transmission probabilities in response to arbitrary patterns of presynaptic inputs.

After fitting the model to real pre- and postsynaptic spike-trains, we compared its behavior to experimentally observed patterns of spike transmission probability. In particular, we compare peaks in the split cross-correlograms to the average model prediction for the same sets of presynaptic spikes (see Methods). We find that our model is flexible enough to explain the changes in synaptic transmission probability observed in spiking statistics for all three synapses above (Fig.

3.4A). Moreover, using the model-based approach, the contributions of each model component can be disentangled. Our results suggest that the pattern of spike transmission probability for the thalamus connection is dominated by a combination of membrane potential summation and short-term depression. Although depression decreases spike transmission probability at shorter ISIs, membrane summation acts to increase postsynaptic spiking. The ANF-SBC synapse, in contrast, shows an increase in spike transmission probability for a medium range of ISIs that is explained by a model dominated by short-term facilitation. Lastly, the VB-Barrel connection shows a higher postsynaptic response for spikes following longer ISIs (isolated) that is explained by the model as an effect of short-term synaptic depression.

In addition to separating the factors affecting spike transmission, the model also improves the prediction of postsynaptic spike timing. To evaluate how spike prediction accuracy is influenced by STP, we compare the prediction of postsynaptic spiking activity after each presynaptic spike for our model with a static model containing all components except STP. In all three datasets, a model with short-term synaptic plasticity provides substantially better predictions of the postsynaptic spiking activity, assessed by Receiver Operating Characteristics (ROC) curves. For the model with short-term synaptic plasticity accuracies were AUC=0.76, 0.70, and 0.79 for the Thalamus pair, VB-Barrel, and ANF-SBC connections, respectively; compared to a model without STP where the model accuracies were AUC=0.54, 0.48, and 0.56.

In our model, STP is described by two coupled differential equations with five parameters: $\theta_{stp} = \{\tau_d, \tau_f, U, f, \tau_s\}$ (see Methods). Here we estimate values for depression, facilitation, and membrane time-constants along with release probability, U, and magnitude of facilitation, f, (Fig. 3.4B). Our result for the thalamus pair shows a higher release probability and depression time-constant with a larger membrane time constant. The VB-Barrel connection shows a higher

depression compared to facilitation time constant with a lower membrane time constant. The ANF-SBC synapse shows a lower release probability compared to the other two connections and a lower depression and membrane time constant. Although here we estimate these parameters from pre- and postsynaptic spiking alone, they could also be estimated from intracellular measurements [33]. We are not aware of any *in vivo* experiments that measured depression or facilitation time-constants for these systems. However, previous *in vitro* studies found a wide range of paired-pulse ratios (0.3 to 0.9) in thalamocortical projections [105], consistent with the depressing VB-Barrel synapse here. Additionally, *in vitro* observations of ANF-SBC connections report depression time-constants on the order of 2-25ms in response to a 100 Hz stimulus train [106,107]. These previous estimates are substantially faster than the time-constants estimated by the TM-GLM for the ANF-SBC connection here. However, as mentioned in [106], different patterns of presynaptic input (e.g. regular, Poisson, natural) can result in different time constants, which makes it difficult to compare *in vitro* and *in vivo* STP parameters directly. One parameter that may be more readily comparable across preparations is the membrane time-constant. We find that the estimated membrane time-constant from the TM-GLM for the thalamus pair is consistent with thalamus relay cells observed intracellularly (12.2 ± 1.1 ms (n=8)) [108], and the estimated membrane time-constant for ANF-SBC is approximately consistent with *in vitro* measurements (1.05 ± 0.09 ms) [106], as well.

Previous work modeling intracellular recordings suggests that the full TM model may not be necessary to explain STP at some, purely depressing synapses [33]. Therefore, we explored how simplified TM models of STP, with fewer parameters, compare with the full model using the Akaike information criterion (AIC; see Methods). AIC evaluates model accuracy (log-likelihood) penalized by the number of parameters and determines if a simplified model with fewer parameters is preferred over a more complex model. We compare the full model to five reduced models: 1) a

model with only integration, without dynamic release probability and resources ($\tau_d, \tau_f = 0$ and $f, U = 1$), 2) a facilitation only model ($\tau_d = 0$), 3) a depression only model ($\tau_f = 0$), 4) a 3-parameter TM model where the magnitude of facilitation is fixed ($f = U$), and 5) a full TM model without resetting integration when a postsynaptic spike occurs ($\pi_i = 1$). For the thalamus pair and VB-Barrel, a model with fixed magnitude of facilitation ($f = U$) performs better while for the ANF-SBC connection the full model gives a better prediction. The full TM model performs well in all cases, but, for some synapses, as previous results suggested [33], there may be ambiguity with parameter identifiability where many parameter settings explain the data.

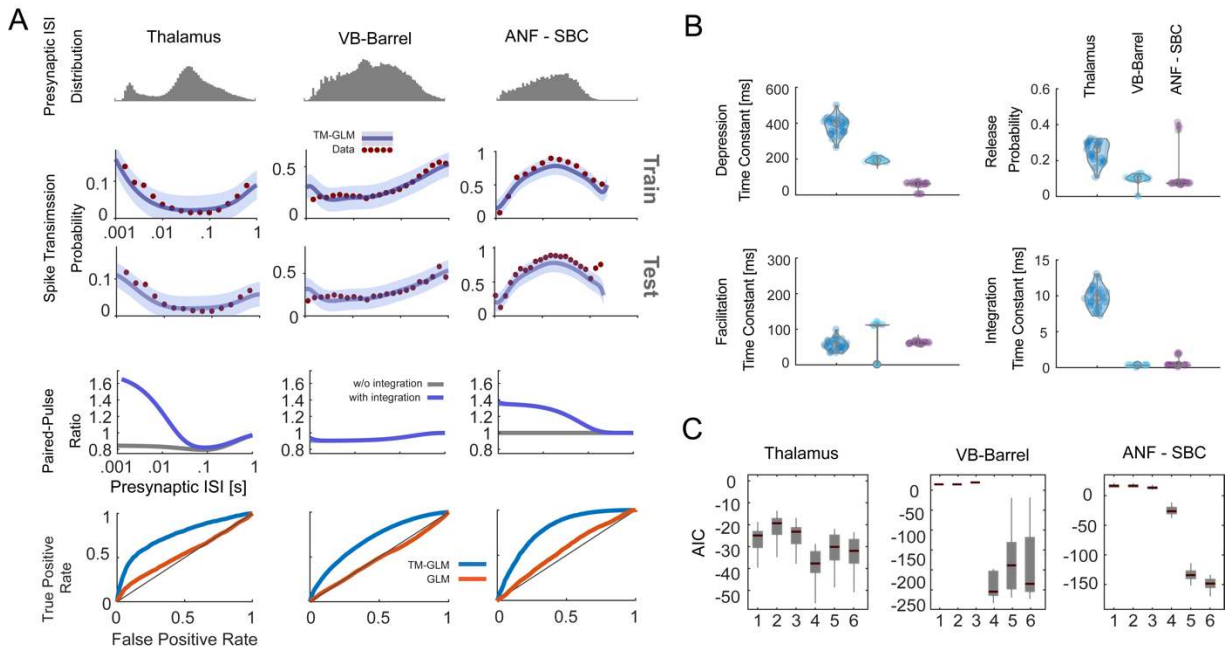

Fig. 3.4: Model predictions of spike transmission dynamics. A: *Spike transmission patterns are diverse across different connections. For three different connections (a pair in thalamus, ventrobasal thalamus to somatosensory cortex, auditory nerve fiber to spherical bushy cell) transmission patterns are explained by a combination of different factors. For each synapse, top panels show the presynaptic ISI distributions (log-spaced). In the second/third row, the observed spike transmission probability (red data points) and model predictions (blue with 95% confidence bands) for training and test set (2-fold cross-validation). We then used the estimated TM parameters for each synapse and simulated responses to for paired presynaptic pulses. Blue curves denote the PPRs of the full model, and gray lines denote PPRs for a model without synaptic summation. In the fourth row, we evaluate how accurately the TM-GLM can prediction individual postsynaptic transmission events. For each individual presynaptic spike, we compare the model transmission probability with the observed binary outcome. ROC curves show the prediction accuracy for the TM-GLM (blue) and a standard GLM without STP (orange).* ***B:*** *Estimates for the four STP parameters of the model for each synapse. Each dot represents estimation from a distinct*

bootstrap sample. ***C:*** *Model comparison for 6 different models (Akaike information criteria relative to a model without plasticity). Models: 1) Integration only ($\tau_d, \tau_f = 0$ and $f, U = 1$), 2) Facilitation only ($\tau_d = 0$), 3) Depression only ($\tau_f = 0$), 4) 3-parameter TM ($f = U$), 5) 4-parameter TM without resetting integration ($\pi_i = 1$), 6) 4-parameter TM.*

Recent patterns of pre- and postsynaptic spiking shape the synaptic transmission probability

Although previous studies have focused largely on how spike transmission probability varies as a function of the single preceding presynaptic ISI, synaptic dynamics depend on the full sequence of presynaptic spiking. Unlike *in vitro* experiments where the state of the synapse can, to some extent, be controlled before studying responses to a specific presynaptic pattern, *in vivo* measurements of spike transmission can be heavily influenced by higher-order correlations between successive ISIs [87]. Additionally, it is difficult to assess the effects of multi-spike patterns empirically by splitting the correlograms, since the number of observations for any given presynaptic spike pattern rapidly decreases with the number of spikes in the pattern. Here we examine how spike transmission depends, not just on the preceding presynaptic ISI, but on triplets of spikes. We compare the empirically observed spike transmission probability following triplets to the estimated spike transmission probability from the TM-GLM. Then, after fitting the TM-GLM, we simulate postsynaptic responses to isolated, local patterns of spikes and determine to what extent the observed spike transmission patterns are influenced by higher-order correlations between successive ISIs.

First, in addition to the timing of the two preceding presynaptic spikes (ISI_1), we split correlograms based on the timing of the three preceding presynaptic spikes using both ISI_1 and ISI_2. Since the TM-GLM provides estimates of the post-synaptic spike probability following every presynaptic spike, we can split both the data and model fits the same way (Fig. 3.5A). We find that the spike

transmission patterns clearly depend on the triplet patterns of presynaptic spikes. That is, the spike transmission probability is influenced by both ISI_1 and ISI_2, and their interaction differed between synapses, as expected from the TM-GLM model. However, similar to the descriptions of spike transmission as a function of ISI_1, the TM-GLM accurately captures the patterns of spike transmission for triplets of presynaptic spikes at those three synapses. In the thalamus pair, spike transmission was dominated by ISI_1, and the effect of ISI_2 appears to be weak or, at least, doesn't appear to be monotonic. Spike transmission at the VB-Barrel connection depends strongly on both ISI_1 and ISI_2, with higher spike transmission probability for longer ISI_2, consistent with recovery from depression. Lastly, for the ANF-SBC connection, transmission probabilities decrease for shorter ISI_2, but there also appears to be a strong interaction between ISI_1 and ISI_2, where transmission probability is high for multiple combinations of these two intervals (e.g. intervals of 10ms then 100ms and intervals of 100ms then 10ms both result in high probability transmission). To examine to what extent the empirical observations of spike transmission are affected by higher-order correlations between successive ISIs, we again use the estimated parameters in the TM-GLM to simulate postsynaptic responses to hypothetical, isolated triplets of presynaptic spikes. In these simulations we fix the post-spike history effect and the excitability in the model to their average values from model fits, and we fix the initial STP state (initial values of R and u in TM model) for the first spike in triplets to the average R and u values from the model fits. In experimental data, the initial state of the pre- and postsynaptic neurons before the triplets occur can wildly differ between different values of ISI_1 and ISI_2. By simulating, we can compare the influence of different triplets (ISI_1 and ISI_2) when the pre- and postsynaptic neurons start at the same state. Here we find that for the thalamus pair, although the empirical data showed no clear effect for ISI_2, the simulated spike transmission probability increases with short ISI_2, consistent with strong synaptic

summation. One reason that this effect may be masked in the empirical transmission probabilities is that post-spike history effects could act to decrease the probability of future postsynaptic spikes. For the VB-Barrel simulations, we find that short ISI_2 decreases transmission probability, consistent with the empirical transmission patterns, although less pronounced. Serial correlations in the sequence of presynaptic spikes (such as long bursts) could act to accentuate the depression in the empirical observations beyond what we see with the simulated responses to isolated triplets. Finally, for the ANF-SBC, although the empirical transmission probability showed decreased transmission for short ISI_2, the simulated responses to isolated patterns have increasing transmission at short ISI_2 (due to synaptic summation). This difference is likely due to the post-spike history effects, which has been fixed for the simulations, but can have a large effect in the experimental data. Since the overall efficacy of this synapse is quite high (>0.7), is likely that a postsynaptic spike follows the first or second presynaptic spike which then influences the response to the third spike.

To better understand the effects of post-spike history, we examined how the postsynaptic spiking history changes the spike transmission patterns with a similar approach. In addition to splitting the correlograms based on ISI_1, we also split based on the previous postsynaptic ISI, ISI_{post} (Fig. 3.5B). Here, as with the triplets of presynaptic spikes, we find that the spike transmission patterns depend on the triplet patterns of 2 pre- and 1 postsynaptic spike, and the TM-GLM accurately captures the patterns of spike transmission at our three synapses (Fig. 3.5B). Here, for both thalamus and VB-Barrel pairs, synaptic transmission probability decreases after a long postsynaptic ISI for all values of ISI_1. In contrast, the ANF-SBC connection shows decreased transmission probability at short postsynaptic ISIs.

As with the triplets of presynaptic spikes, we then simulate how these local patterns of 2 pre- and 1 postsynaptic spike change spike transmission probability when the neurons start from the same initial conditions (average values of excitability, post-spike history, R and u). For the thalamus and VB-Barrel pairs, the simulations of isolated, local patterns match the general trends of empirical spike transmission. However, for the VB-Barrel synapse, the effect of ISI_{post} in the empirical transmission patterns is stronger than in the simulations, suggesting that serial correlations in ISIs could again play a role and does decrease transmission probability for isolated patterns. However, as with the responses to triplets of presynaptic spikes, these local patterns alone are insufficient to explain the empirically observed patterns of spike transmission.

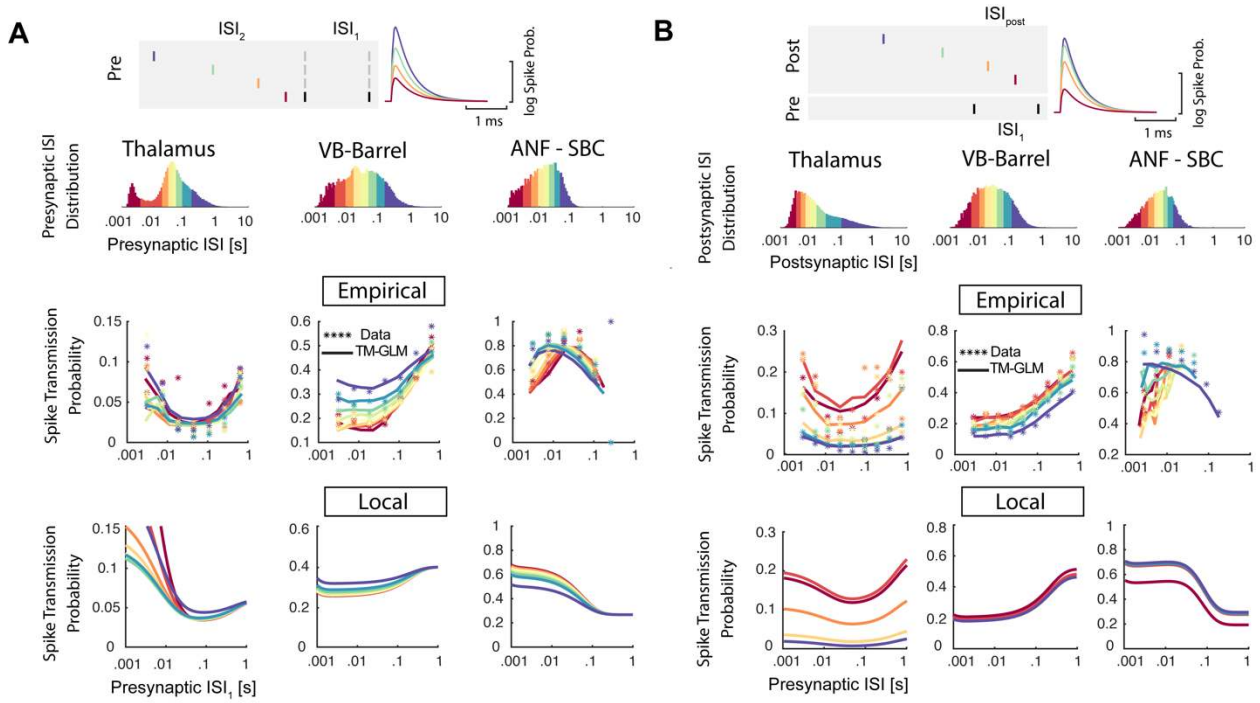

Fig. 3.5: Effects of pre- and postsynaptic spiking patterns. A: *Synaptic transmission patterns change based on the full sequence of presynaptic spiking. Top panel shows a schematic of 4 different patterns of presynaptic spike triplets with a fixed ISI between the two most recent presynaptic spikes (black dashed lines). We split the presynaptic ISI distribution into 8 quantiles. Each data point shows the observed spike transmission probability corresponding to the* ISI_2 *quantile with the same color. Solid lines are the average estimated probability for each pattern under the model (based on the full sequence of observed spikes). To examine the influence of serial correlations, we then stimulate model responses to the isolated triplet pattern, assuming the synapse is initially in an average state.* ***B:*** *Synaptic transmission patterns change depending on the history of postsynaptic spiking, as well. Here each data point in the scatter plots shows the spike transmission probability of the corresponding to the postsynaptic ISIs of the same color in the ISI distribution. Colors represent the corresponding time difference between presynaptic and previous postsynaptic spike. Solid lines are the average predicted probability for quantiles with*

corresponding colors. Last row shows simulations of the model using estimated STP parameters and fixing the excitability from the model fits to their average values. Here the history effect for each ISI interval is set to the post-spike history filter value on that interval.

Spike transmission patterns change depending on stimulus type

These results suggest that the presynaptic spike pattern has a complex effect on spike transmission probability. In sensory systems, one variable that affects the presynaptic spike pattern is the external stimulus. To examine how differences in stimulus statistics might alter spike transmission, we fitted our model to a dataset recorded juxtacellularly from an ANF-SBC synapse, presented with *natural sounds*, a range of randomized frequency-level pure-tones (*tuning stimuli*), and s*pontaneous activity* in the absence of acoustic stimulation. We merged these three datasets and fitted the model to the merged dataset. As with the previous fits of the ANF-SBC connection (based on a different set of *tuning stimuli*), the transmission probability under all three conditions exhibits a bandpass-like pattern suggesting facilitation and little to no synaptic summation. However, spike transmission during natural stimuli was markedly different from that during pure tone stimulation. During natural sounds, transmission probability is maximized at 100 ms rather than 10 ms in the *tuning stimuli* and during *spontaneous activity*. Further, *natural stimuli* have much lower transmission probability at short ISIs. Interestingly, the TM-GLM captures the overall facilitation, but also captures differences due to the different stimuli. In contrast, a static GLM captures almost none of the variations in spike transmission probability suggesting that a fixed coupling term, postsynaptic history, and, slow fluctuations of presynaptic spiking are not sufficient to capture patterns of spike transmission probabilities (Fig. 3.6A). Together, these results suggest that the combination of STP, synaptic summation, history, and excitability is sufficient to explain the observed differences between stimuli, without requiring any additional adaptation or plasticity.

Since these recordings were performed juxtacellularly, we also have access to the slope of individual (extracellular) PSPs, which correlates with the intracellular PSP amplitude. We compared patterns of individual PSP slopes for each stimulus type and how they correlate with the estimated coupling amplitude following individual presynaptic spikes in our model (Fig. 3.6B, 5C). Note that patterns of PSP slopes do not have the same pattern as spike transmission probability, since there are other factors (e.g. postsynaptic spiking history) contributing to postsynaptic spiking. These results show the stimulus-dependence of PSP amplitudes and a static GLM without STP cannot account for these variations. Although the correlation is not perfect, the model does correlate with the measured PSP slope, even though the model only has access to spikes. By modeling dynamic functional connectivity, we can approximately reconstruct the amplitude of individual synaptic events.

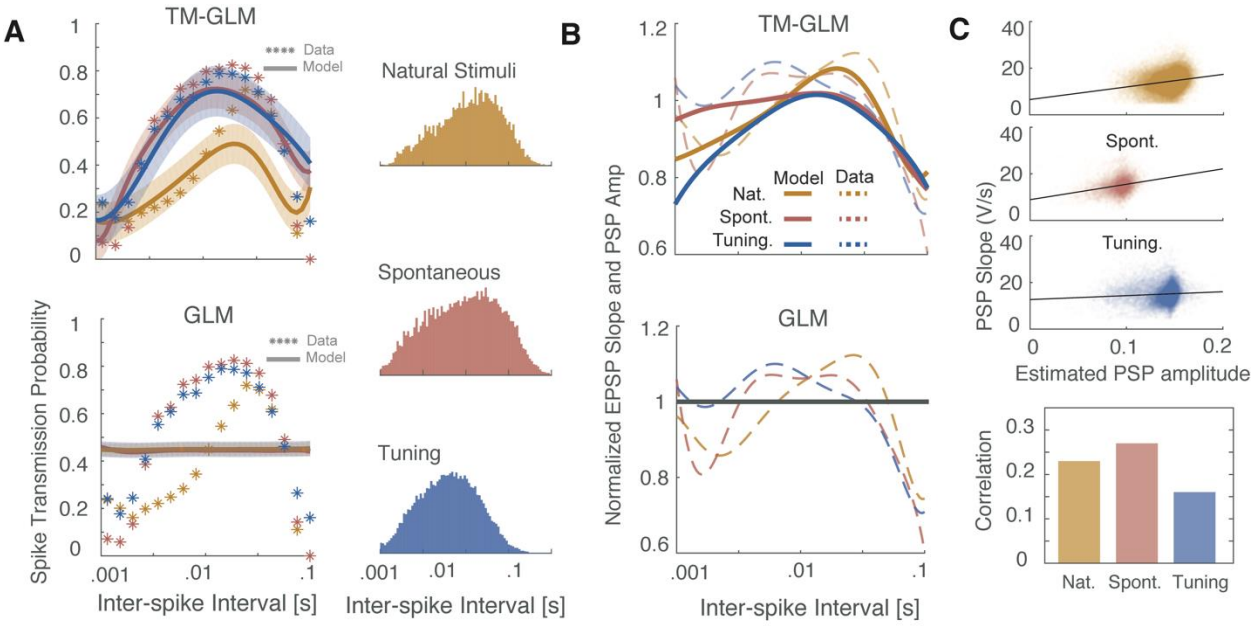

Fig. 3.6: The TM-GLM captures stimulus-dependent changes in spike transmission probability at the ANF-SBC synapse. A) *The TM-GLM captures stimulus-dependent spike transmission probability patterns better than a static model without short-term synaptic plasticity. Asterisks show spike transmission probability for (log-spaced) presynaptic ISIs during two types of auditory stimuli and during spontaneous activity: Natural Sounds (yellow), Spontaneous Activity (red), and Tuning Stimuli (blue). Solid lines and 95% confidence bands show model predictions for each stimulus type. Corresponding inter-spike interval distributions are shown on the right.* ***B)*** *The TM-GLM captures changes in extracellularly recorded PSPs. Here the observed PSP slope (dashed lines) approximately matches and coupling term in the TM-GLM (solid lines) for each three stimuli. Although the spike transmission probability of the static GLM can vary as a function of presynaptic ISI due to non-synaptic factors, the coupling term is fixed.* ***C)*** *Estimates of individual PSP amplitudes predicted by the model and their PSP slopes in the juxtacellular recording. Black lines denote linear fits and the bar plot shows the corresponding Spearman correlations.*

Postsynaptic cell-type specific changes in spike transmission patterns

We applied our model to spiking data from a large-scale multi-electrode array recording to investigate the spike transmission dynamics in synapses from putative excitatory neurons to two different putative inhibitory subtypes. We detected putative synapses using the log-likelihood ratio (LLR < -6, ~200 synapses) between a full model of the correlogram that includes the synaptic effect and smooth model of the correlogram that only captures the slow structure (see Methods). We then found excitatory-inhibitory microcircuits where putative excitatory neurons (based on the cross-correlogram and spike waveform) give inputs to putative inhibitory neurons (41 excitatory synapses onto 9 inhibitory neurons in total). To identify inhibitory neurons as inhibitory, we required the neuron to have an outgoing connection to a third neuron with a fast, transient decrease in the cross-correlogram. Each of the 9 putative inhibitory neurons here had at least one outgoing connection where the spiking probability of a downstream neuron decreases >18% relative to baseline following its spiking (Fig. 3.7A). We then categorized each neuron as a putative fast-spiking (FS, n=5) or regular-spiking (RS, n=4) unit based on the spike waveform and firing rate (Fig. 3.7B). Putative FS units had narrow-width spike waveforms (half-width of the trough = .08$\pm$.02 ms) and higher firing rates (26.07$\pm$9.6 Hz) compared to putative RS neurons (n=4) with broader waveforms (half-width = .14$\pm$.02 ms) and lower firing rate (10.18$\pm$10.01 Hz).

We identified these microcircuits in different regions with 4 putative excitatory-inhibitory microcircuits recorded in hippocampus (depth differences: 77.2$\pm$49.4 μm), 3 in thalamus (49.4$\pm$26.2 μm), and 2 in motor cortex (36.4$\pm$23.5 μm). Putative excitatory neurons showed a wide spike waveform (half-width = .18$\pm$.04 ms) similar to the putative regular-spiking inhibitory neurons, but these two classes can be distinguished by their outgoing connection types (e.g. inhibitory/excitatory) [109] (Fig. 3.6B). Average efficacies from putative excitatory-FS

connections (.22±.12, n=22) were larger, on average, compared to putative excitatory-RS efficacies (.13±.13, n=19). We then fit the TM-GLM to data from these 41 putative synapses, similar to the three identified synapses above. We find that the STP parameters for these two types of synapses largely overlap, except for the membrane time-constant. Interestingly, the membrane time-constants measured for these inhibitory subtypes *in vitro* overlap with our estimates here (Fig. 3.7C) [110]. Although *in vitro* studies have not explored the same TM model used here, there is evidence of postsynaptic cell-type specific STP where putative excitatory-RS connections show facilitation and putative excitatory-FS connections show depression [71]. Here we find that both connection types are somewhat facilitating but excitatory-FS connections having a slightly shorter facilitation time constant. However, unlike what would be expected if excitatory-FS connections were depressing, the release probability of excitatory-FS connections is lower than for excitatory-RS connections (Fig. 3.7C).

To better understand synaptic transmission *in vivo* it is important to consider not just the parameters of the synapse but the full history of presynaptic spiking in the individual presynaptic neurons. We use the estimated model parameters to simulate responses to a train of regular presynaptic spikes with the frequency matched to the average firing rate of the corresponding excitatory input. In simulating postsynaptic responses to the spike train, we fix the excitability and postsynaptic history to their average values from model fits and set the initial STP state of the first spike in the train to the average R and u values from model fits. With these input-matched simulations, excitatory-RS connections show a higher postsynaptic potential compared to excitatory-FS connections (Fig. 3.7D). Similarly, we simulated the paired-pulse ratio (PPR) at different inter-stimulus intervals in our TM model following the average state. On average, connections to regular-spiking inhibitory neurons show a higher PPR (Fig. 3.7E). For all

connections, we then evaluated the spike prediction accuracy of a model without STP (e.g. static GLM) with our TM-GLM using the Area Under the ROC Curve (Fig. 3.7F). The model with STP (TM-GLM) gives more accurate predictions when the postsynaptic neuron spikes following a presynaptic spike for our population of 41 putative excitatory-inhibitory connections (AUC=.69±.05) in comparison with the static GLM (AUC=.50±.03). Altogether, these results illustrate how a dynamic model of functional connectivity, such as the TM-GLM, may allow us to investigate cell-type-specific differences in short-term synaptic dynamics in behaving animals using only pre- and postsynaptic spiking.

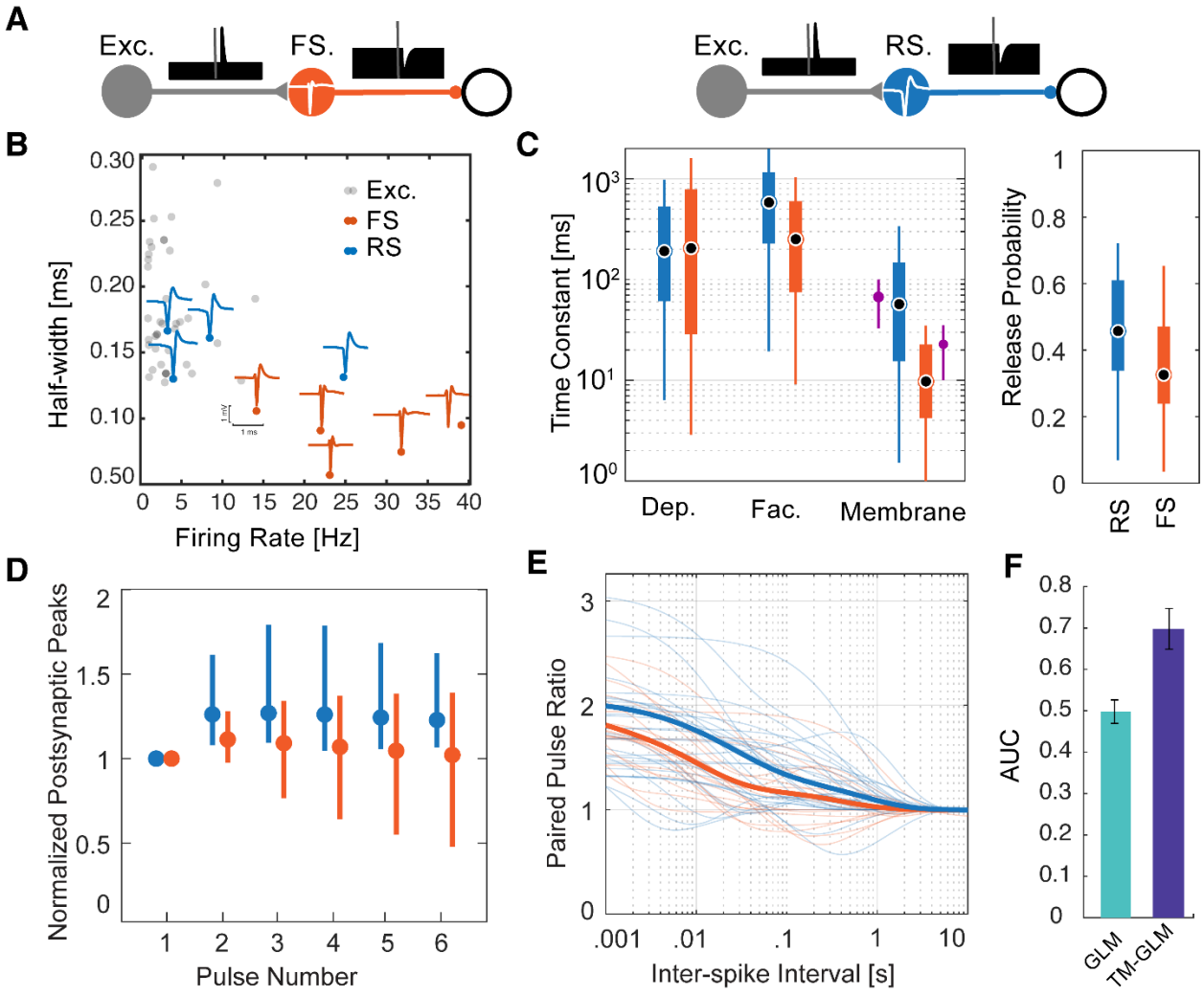

Fig. 3.7: Distinctive short-term synaptic plasticity dynamics in connections between excitatory neurons to putative Regular-Spiking (RS) and Fast-Spiking (FS) inhibitory neurons. *A) Here we examine putative synapses between excitatory neurons and inhibitory neurons (identified by their cross-correlations) and separate the putative inhibitory neurons into two classes: fast-spiking, which have narrow spike waveforms and high rates (left), and regular-spiking (right), which have wide waveforms and lower rates. Identifying these synapses requires both finding both a putative excitatory input and a putative inhibitory output for the same neuron. B) Half-widths (of the trough) of the spike waveforms and firing rates for the FS (orange) and RS (blue) inhibitory neurons, as well as, their excitatory inputs (grey). Individual blue and orange waveforms*

(maximum amplitude across the MEA) are shown for all 9 putative inhibitory neurons. C) Estimated depression, facilitation, and membrane time-constants for excitatory-RS and excitatory-FS connections, along with the release probability (right). The purple error-bar next to the membrane time-constant estimations show the median and standard deviations from in vitro experiments [110]. *D) Simulated postsynaptic potential amplitudes estimated from Tsodyks-Markram model of short-term synaptic plasticity using estimated parameters. For each synapse, PSPs are estimated in response to a pulse train with inter-pulse intervals set to their corresponding average presynaptic inter-spike intervals. Dots and error bars denote the median and inter-quartile range for excitatory-RS (blue) and excitatory-FS (red) connections. E) Simulated Paired-Pulse Ratio for individual synapses of excitatory-RS (blue) and excitatory-FS (red) connections as a function of the presynaptic ISI. F) Area Under the Curve (AUC) of postsynaptic spiking prediction using the static GLM without short-term synaptic plasticity (green) and the TM-GLM with short-term synaptic plasticity (blue).*

Discussion

Short-term synaptic plasticity (STP) has been extensively studied *in vitro* and with intracellular recordings where the amplitudes of individual postsynaptic potential/currents (PSP/PSCs) can be directly measured. By using controlled experiments with specific, structured presynaptic spike patterns these studies established how short-term synaptic dynamics can be described by the interactions between release probability and vesicles/resource dynamics [11]. These alterations in PSP/PSP amplitudes can affect the statistics of postsynaptic spiking. Thus, STP could, explain why the probability of postsynaptic spiking depends not just on the presence of a presynaptic spike, but on the timing of the most recent presynaptic inter-spike interval [20,22]. However, the relationship between STP and *in vivo* spike transmission patterns is complex. Patterns of postsynaptic spike transmission are highly diverse and multiple factors beyond STP and the most recent presynaptic ISI shape these patterns. Here we aimed to disentangle the different contributions to spike transmission by developing an augmented generalized linear model, the TM-GLM that explicitly includes STP dynamics as slow changes in postsynaptic excitability and the history of postsynaptic spiking.

Synapses with different types of STP can allow the same sequence of presynaptic spikes to generate different patterns of postsynaptic spiking and thereby control the information flow in the brain. Here we tracked the observed spike transmission probability of three strong synapses from different species and brain areas. The dynamical spike transmission model enables us to disentangle different factors (e.g. slow firing rate changes, postsynaptic spiking history, synaptic summation, and STP) that shape these diverse patterns. First, we investigate the role of STP and the full sequence of presynaptic spiking activity in shaping the spike transmission patterns. In three strong synapses (an intra-thalamic synapse, a thalamocortical synapse, and an auditory brainstem

synapse) we show how models of functional connectivity with short-term synaptic plasticity can 1) capture diverse pattern of spike transmission probability, 2) disentangle these transmission patterns to the multiple factors that shape postsynaptic response, 3) extract biologically plausible synaptic dynamics, and 4) improve prediction of postsynaptic spiking.

Estimating static functional connectivity using spike times has revealed network structure in the retina [24] and hippocampus [48], can reconstruct true physiological circuitry [25], and improves encoding and decoding [24,49,92]. However, synaptic weights change over a wide range of timescales depend on external stimuli and behavior [81]. Additionally, synaptic dynamics can shape information transmission in different ways for different pattern of presynaptic spiking, e.g. different behavioral tasks. Although, standard GLMs can partially capture the first-order effects of recent presynaptic spikes on postsynaptic spiking probability, they fail to capture the nonlinear dynamics of synaptic transmission affected by the whole sequence. Here, in a recording from the endbulb of Held (ANF-SBC) we found that spike transmission patterns differed for different stimuli (natural sound stimuli, varying pure tones and without stimulation - e.g. spontaneous activity), and these differences were well-described by the TM-GLM. Although the STP-parameters were the same for all stimuli, the different presynaptic spike patterns yield different synaptic dynamics and different patterns of spike transmission. Since spike transmission probability in the TM-GLM depends on the full history of presynaptic spiking, this model can account for changes on behavioral timescales even in the absence of adaptation or other forms of plasticity (e.g. STDP, LTP).

Cell-type specific interactions in layers and regions of the brain perform different computational tasks. Previous *in vitro* studies have shown that STP dynamics depend on both presynaptic and postsynaptic cell-types [71]. Here in a large multi-electrode array recording of a freely behaving

mouse we investigated STP dynamics of synaptic connections from putative excitatory neurons to two different subtypes of putative inhibitory neurons: fast-spiking and regular-spiking. Using inferred short-term dynamics, predicted responses to train of spikes with the same input frequencies as the presynaptic neuron in those connections show facilitation in excitatory-RS and depression in excitatory-FS connections which are in line with previous *in vitro* findings [71]. Moreover, the model with short-term dynamics significantly improves the prediction of postsynaptic activity. As large-scale extracellular recordings advance, models such as the TM-GLM are promising to characterize and compare the short-term dynamics of spike transmission of many different cell types, brain regions, and species.

Although our model provides a tool to characterize the dynamics of spike transmission, there are limitations on how well TM-GLM can capture true synaptic dynamics. Firstly, functional connections inferred from spikes do not necessarily guarantee anatomical connections. A peak in the cross-correlogram does not uniquely indicate the presence of a monosynaptic connection [100,111]. Here we assume that the transient, short-latency increase in postsynaptic spiking activity following a presynaptic spike indicates the presence of an excitatory monosynaptic connection [19]. Nevertheless, verifying connections using optogenetics, juxtacellular recordings [112], imaging [113] or anatomical reconstruction provide a more accurate estimate of true anatomical connections. Secondly, to model short-term dynamics in spiking neurons we employ a rate model that does not explicitly account for the detailed membrane potential of the postsynaptic neuron. Other approaches to modeling synaptic transmission with realistic spike-generation mechanisms, currents, and even dendritic morphology do exist, but are typically less computationally tractable [114]. Here we employed an augmented GLM with a logistic spike nonlinearity. We chose the logistic nonlinearity over the conventional exponential function as it

appears to better describe strong connections, such as the ANF-SBC, but other nonlinearities may be better for other neurons [51]. There are also alternatives to the Tsodyks-Markram model for modeling synaptic dynamics. The TM model is biologically plausible, but, since it is deterministic, it ignores the stochasticity of synaptic release and only tracks the dynamics of average postsynaptic potentials. Finally, there are many covariates that could be added to improve model performance, including local field potentials, connections to other simultaneously observed presynaptic neurons [92], higher-order history or coupling terms [115,116], and covariates related to other types of plasticity [13,67,115,117,118]. Despite these simplifying assumptions and the fact that we only observe a fraction of inputs to the neuron, the TM-GLM captures a wide diversity of *in vivo*, excitatory spike transmission patterns.

Short-term synaptic plasticity alters information transmission from presynaptic to postsynaptic neurons by dynamically changing the synaptic efficacy [11,20,22]. Intracellular studies *in vitro* or with artificial stimulation patterns have shown that short-term synaptic dynamics depend on cell-types and brain regions [36,71]. However, there is evidence that, in addition to these anatomical dependencies, short-term synaptic dynamics also depend on stimulus type and the larger computational function of the neural circuit [119]. To understand how these synaptic dynamics alter neural computations we will need to study them during natural patterns of presynaptic spiking [6] and, ultimately, during natural behavior. Since large-scale intracellular recordings are currently not feasible *in vivo*, here we examined the possibility of using existing large-scale extracellular recordings to quantify the dynamics of spike transmission and infer the short-term dynamics of synaptic responses. We find that including STP in models of spiking neurons can capture diverse patterns of spike transmission, including patterns that are stimulus-dependent and cell-type-specific. Additionally, these models substantially improve prediction of postsynaptic spiking

following presynaptic spikes and, at least in some cases, can approximately reconstruct individual PSP amplitudes.

Acknowledgements

Thanks to Nick Steinmetz for providing MEA dataset. AG, NR, and IHS were supported by NSF CAREER 1651396 to IHS.

Chapter 4: Discussion

Short-term synaptic plasticity (STP) has traditionally been studied using intracellular, *in vitro* recording techniques where the membrane potential of a postsynaptic neuron can be directly observed from the postsynaptic potential (PSP). By measuring changes in PSP over time such intracellular, *in vitro* experiments have characterized many different forms of plasticity and identified the underlying molecular mechanisms and demonstrated how STP differs across cell-types [2,68]. At specific synapses, STP can also change during development, as neuromodulator concentrations change, and after long-term plastic changes. Typically, *in vitro* experiments characterize STP using periodic pulses with fixed intervals. However, extending findings from *in vitro* experiments to understand how synapses change *in vivo* and during behavior is a major challenge. Measuring how synapses between neurons change over time is experimentally difficult and typically requires precise intracellular recordings and electrical stimulation. However, advances in neural recording technology are providing increasingly large datasets where the extracellular spiking activity of hundreds of neurons can be measured simultaneously. Our goal here was to build on static models of functional connectivity from spike observations and develop a framework that allows interactions between neurons to be dynamic. Here we developed methods that allow time-varying synaptic weights to be inferred from spike times alone. These models pave a way to study short-term synaptic plasticity and network dynamics that are crucial for temporal information processing in behaving animal.

In chapter 2, we introduced two methods (eTM-GLM and GBLM) for estimating short-term synaptic plasticity from spikes alone. To validate these methods, we used an *in vitro* current injection experiment where an artificial population of presynaptic neurons – whose spiking and plasticity parameters are known – acted as input to an intracellularly recorded neuron. Both models were able to robustly characterize different plasticity types and reconstructed PSP amplitudes. The advantage of these models is the ability to infer time-varying synaptic input using natural, ongoing spike activity.

In chapter 3, we built upon eTM-GLM in the previous chapter and examined short-term synaptic dynamics *in vivo*. Here, we used a model-based approach (an extended version of a GLM; TM-GLM) to incorporate the influences of membrane potential integration, post-spike history, and postsynaptic excitability. By incorporating all of these factors, we provided a detailed explanation for the observed patterns of spike transmission and also made accurate predictions about the postsynaptic responses to individual presynaptic spikes.

Here we have shown that dynamic functional connectivity can describe synaptic transmission properties in short timescales. This work in its current state could be a useful tool for systems and computational neuroscientists interested in studying synaptic transmission *in vivo*. Moreover, future methods could improve the current work by including analytical description of short-term synaptic plasticity that guarantees identifiability. Also, this work shows how synaptic transmission and neural computation are task and cell-type specific. In future work, this insight could be incorporated in optimizing generative models of spiking neural networks.

References

1. Zucker RS, Regehr WG. Short-Term Synaptic Plasticity. Annu Rev Physiol. 2002;64: 355–405. doi:10.1146/annurev.physiol.64.092501.114547
2. Abbott LF, Nelson SB. Synaptic plasticity: taming the beast. Nat Neurosci. 2000;3: 1178–1183. doi:10.1038/81453
3. Whitlock JR, Heynen AJ, Shuler MG, Bear MF. Learning Induces Long Term Potentiation in the Hippocampus. Science (80-). 2006;313: 1093–1097. doi:10.1126/science.1128134
4. Bliss T V, Collingridge GL. A synaptic model of memory: long-term potentiation in the hippocampus. Nature. 1993;361: 31–39. doi:10.1038/361031a0
5. Murayama M, Pérez-Garci E, Nevian T, Bock T, Senn W, Larkum ME. Dendritic encoding of sensory stimuli controlled by deep cortical interneurons. Nature. 2009;457: 1137–1141. doi:10.1038/nature07663
6. Klyachko VA, Stevens CF. Excitatory and feed-forward inhibitory hippocampal synapses work synergistically as an adaptive filter of natural spike trains. PLoS Biol. 2006;4: 1187–1200. doi:10.1371/journal.pbio.0040207
7. Lindner B, Gangloff D, Longtin A, Lewis JE. Broadband Coding with Dynamic Synapses. J Neurosci. 2009;29: 2076–2087. doi:10.1523/JNEUROSCI.3702-08.2009
8. Mongillo G, Barak O, Tsodyks M. Synaptic theory of working memory. Science. 2008;319: 1543–1546. doi:10.1126/science.1150769
9. Hiratani N, Fukai T. Interplay between short- and long-term plasticity in cell-assembly formation. PLoS One. 2014;9. doi:10.1371/journal.pone.0101535
10. Tsodyks M, Markram H. The neural code between neocortical pyramidal neurons depends. Proc Nat Acad Sci USA. 1997;94: 719–723.
11. Markram H, Wang Y, Tsodyks M. Differential signaling via the same axon of neocortical pyramidal neurons. Proc Natl Acad Sci U S A. 1998;95: 5323–8. doi:10.1073/pnas.95.9.5323
12. Dayan P, Abbott LF. Theoretical Neuroscience [Internet]. Computational and Mathematical Modeling of Neural …. 2001. doi:10.1016/j.neuron.2008.10.019
13. Linderman S, Stock C, Adams R. A framework for studying synaptic plasticity with neural spike train data. Adv Neural Inf. 2014; Available: http://arxiv.org/abs/1411.4077
14. Hennig MH. Theoretical models of synaptic short term plasticity. Front Comput Neurosci.

2013;7: 45. doi:10.3389/fncom.2013.00045

15. Froemke RC, Dan Y. Spike-timing-dependent synaptic modification induced by natural spike trains. Nature. Nature Publishing Group; 2002;416: 433–8. doi:10.1038/416433a
16. Turrigiano GG, Nelson SB. Rate , Timing , and Cooperativity Jointly Determine Cortical Synaptic Plasticity. Neuron. 2001;32: 1149–1164.
17. Stevenson IH, Kording KP. How advances in neural recording affect data analysis. Nat Neurosci. Nature Publishing Group; 2011;14: 139–142. doi:10.1038/nn.2731
18. Fetz EE, Gustafsson B. Relation between shapes of post-synaptic potentials and changes in firing probability of cat motoneurones. J Physiol. 1983;341: 387–410. doi:10.1113/jphysiol.1983.sp014812
19. Perkel DH, Gerstein GL, Moore GP. Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. Biophys J. 1967;7: 419–440. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1368069/
20. Swadlow HA, Gusev AG. The impact of "bursting" thalamic impulses at a neocortical synapse. Nat Neurosci. 2001;4: 402–8. doi:10.1038/86054
21. Swadlow HA. Thalamocortical control of feed-forward inhibition in awake somatosensory "barrel" cortex. Philos Trans R Soc Lond B Biol Sci. 2002;357: 1717–27. doi:10.1098/rstb.2002.1156
22. English DF, McKenzie S, Evans T, Kim K, Yoon E, Buzsáki G. Pyramidal Cell-Interneuron Circuit Architecture and Dynamics in Hippocampal Networks. Neuron. 2017;96: 505–520. doi:10.1016/j.neuron.2017.09.033
23. Stevenson IH, Rebesco JM, Miller LE, Körding KP. Inferring functional connections between neurons. Curr Opin Neurobiol. 2008;18: 582–588. doi:10.1016/j.conb.2008.11.005
24. Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, et al. Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature. Macmillan Publishers Limited. All rights reserved; 2008;454: 995–9. doi:10.1038/nature07140
25. Gerhard F, Kispersky T, Gutierrez GJ, Marder E, Kramer M, Eden U. Successful Reconstruction of a Physiological Circuit with Known Connectivity from Spiking Activity Alone. PLoS Comput Biol. 2013;9. doi:10.1371/journal.pcbi.1003138
26. Volgushev M, Ilin V, Stevenson IH. Identifying and Tracking Simulated Synaptic Inputs from Neuronal Firing: Insights from In Vitro Experiments. PLoS Comput Biol. 2015;11.

doi:10.1371/journal.pcbi.1004167

27. van Rossum MC, Bi GQ, Turrigiano GG. Stable Hebbian learning from spike timing-dependent plasticity. J Neurosci. 2000;20: 8812–8821. doi:20/23/8812 [pii]
28. Volgushev M, Balaban P, Chistiakova M, Eysel UT. Retrograde signalling with nitric oxide at neocortical synapses. Eur J Neurosci. 2000;12: 4255–4267. doi:10.1046/j.0953-816X.2000.01322.x
29. Volgushev M, Chen J-Y, Ilin V, Goz R, Chistiakova M, Bazhenov M. Partial Breakdown of Input Specificity of STDP at Individual Synapses Promotes New Learning. J Neurosci. 2016;36: 8842–8855. doi:10.1523/JNEUROSCI.0552-16.2016
30. Hardingham NR, Hardingham GE, Fox KD, Jack JJB. Presynaptic efficacy directs normalization of synaptic strength in layer 2/3 rat neocortex after paired activity. J Neurophysiol. 2007;97: 2965–2975. doi:10.1152/jn.01352.2006
31. Markram H, Tsodyks M. Redistribution of synaptic efficacy between neocortical pyramidal neurons [Internet]. Nature. 1996. pp. 807–810. doi:10.1038/382807a0
32. Tsodyks M, Pawelzik K, Markram H. Neural networks with dynamic synapses. Neural Comput 1998 May;10(4)821-835. 1998;835: 821–835. doi:10.1162/089976698300017502
33. Costa RP, Sjöström PJ, van Rossum MCW. Probabilistic inference of short-term synaptic plasticity in neocortical microcircuits. Front Comput Neurosci. 2013;7: 75. doi:10.3389/fncom.2013.00075
34. Varela JA, Sen K, Gibson J, Fost J, Abbott LF, Nelson SB. A quantitative description of short-term plasticity at excitatory synapses in layer⅔ of rat primary visual cortex. J Neurosci. 1997;17: 7926–7940. Available: http://www.jneurosci.org/content/17/20/7926.short
35. Wang Y, Markram H, Goodman PH, Berger TK, Ma J, Goldman-Rakic PS. Heterogeneity in the pyramidal network of the medial prefrontal cortex. Nat Neurosci. 2006;9: 534–542. doi:10.1038/nn1670
36. Markram H, Muller E, Ramaswamy S, Reimann MW, Abdellah M, Sanchez CA, et al. Reconstruction and Simulation of Neocortical Microcircuitry. Cell. 2015;163: 456–492. doi:10.1016/j.cell.2015.09.029
37. Beck O, Chistiakova M, Obermayer K, Volgushev M. Adaptation at synaptic connections to layer 2/3 pyramidal cells in rat visual cortex. J Neurophysiol. 2005;94: 363–376.

Available: http://jn.physiology.org/content/94/1/363.short

38. Li H, Bandrowski AE, Prince DA. Cortical Injury Affects Short-Term Plasticity of Evoked Excitatory Synaptic Currents Cortical Injury Affects Short-Term Plasticity of Evoked Excitatory Synaptic Currents. J Neurophysiol. 2005;93: 146–156. doi:10.1152/jn.00665.2004
39. Volgushev M. Precise Long-Range Synchronization of Activity and Silence in Neocortical Neurons during Slow-Wave Sleep. J Neurosci. 2006;26: 5665–5672. doi:10.1523/JNEUROSCI.0279-06.2006
40. Chauvette S, Volgushev M, Timofeev I. Origin of active states in local neocortical networks during slow sleep oscillation. Cereb Cortex. 2010;20: 2660–2674. doi:10.1093/cercor/bhq009
41. Sedigh-Sarvestani M, Vigeland L. Intracellular, In Vivo, Dynamics of Thalamocortical Synapses in Visual Cortex. J. 2017; Available: http://www.jneurosci.org/content/37/21/5250.abstract
42. Buzsáki G. Large-scale recording of neuronal ensembles. Nat Neurosci. 2004;7: 446–51. doi:10.1038/nn1233
43. Kelly RC, Smith M a, Samonds JM, Kohn A, Bonds a B, Movshon JA, et al. Comparison of recordings from microelectrode arrays and single electrodes in the visual cortex. J Neurosci. 2007;27: 261–264. doi:10.1523/JNEUROSCI.4906-06.2007
44. Usrey WM, Alonso JM, Reid RC. Synaptic interactions between thalamic inputs to simple cells in cat visual cortex. J Neurosci. 2000;20: 5461–5467. doi:20/14/5461 [pii]
45. Paninski L, Pillow J, Lewi J. Statistical models for neural encoding, decoding, and optimal stimulus design. Prog Brain Res. 2007;165: 493–507. doi:10.1016/S0079-6123(06)65031-0
46. Paninski L. Maximum likelihood estimation of cascade point-process neural encoding models. Netw Comput Neural Syst. 2004;15: 243–262. doi:10.1088/0954-898X/15/4/002
47. Brillinger DR. Maximum likelihood analysis of spike trains of interacting nerve cells. Biol Cybern. 1988;59: 189–200. doi:10.1007/BF00318010
48. Harris K, Csicsvari J, Hirase H, Dragoi G, Buzsáki G. Organization of cell assemblies in the hippocampus. Nature. 2003;424: 552–556. doi:10.1038/nature01765.1.
49. Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN, John P. A Point Process

Framework for Relating Neural Spiking Activity to Spiking History, Neural Ensemble, and Extrinsic Covariate Effects. J Neurophysiol. 2005;93: 1074–1089. doi:10.1152/jn.00697.2004

50. Ahrens MB, Linden JF, Sahani M. Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. J Neurosci. 2008;28: 1929–1942. doi:10.1523/JNEUROSCI.3377-07.2008
51. McFarland JM, Cui Y, Butts DA. Inferring Nonlinear Neuronal Computation Based on Physiologically Plausible Inputs. PLoS Comput Biol. 2013;9. doi:10.1371/journal.pcbi.1003143
52. Latimer KW, Chichilnisky EJ, Rieke F, Pillow JW. Inferring synaptic conductances from spike trains with a biophysically inspired point process model. Neural Inf Process Syst. 2014;27: 954–962. Available: http://papers.nips.cc/paper/5262-inferring-synaptic-conductances-from-spike-trains-with-a-biophysically-inspired-point-process-model
53. Bird AD, Wall MJ, Richardson MJE. Bayesian Inference of Synaptic Quantal Parameters from Correlated Vesicle Release. Front Comput Neurosci. 2016;10: 116. doi:10.3389/fncom.2016.00116
54. Schmidt M. minConf - projection methods for optimization with simple con- straints in matlab [Internet]. 2008. Available: http://www.di.ens.fr/ mschmidt/Software/minConf.html
55. Ilin V, Stevenson IH, Volgushev M. Injection of fully-defined signal mixtures: A novel high-throughput tool to study neuronal encoding and computations. PLoS One. 2014;9. doi:10.1371/journal.pone.0109928
56. Liu YH, Wang XJ. Spike-frequency adaptation of a generalized leaky integrate-and-fire model neuron. J Comput Neurosci. 2001;10: 25–45. doi:10.1023/A:1008916026143
57. Barri A, Wang Y, Hansel D, Mongillo G. Quantifying Repetitive Transmission at Chemical Synapses: A Generative-Model Approach. eNeuro. 2016;3: Suppl:1-40. doi:10.1523/ENEURO.0113-15.2016
58. Ventura V. Traditional waveform based spike sorting yields biased rate code estimates. Proc Natl Acad Sci U S A. 2009;106: 6921–6. doi:10.1073/pnas.0901771106
59. Testa-Silva G, Verhoog MB, Linaro D, de Kock CPJ, Baayen JC, Meredith RM, et al. High Bandwidth Synaptic Communication and Frequency Tracking in Human Neocortex. PLoS Biol. 2014;12. doi:10.1371/journal.pbio.1002007

60. Nadim F, Bucher D. Neuromodulation of neurons and synapses. Curr Opin Neurobiol. 2014;29: 48–56. doi:10.1016/j.conb.2014.05.003
61. Kulkarni JE, Paninski L. Common-input models for multiple neural spike-train data. Network. 2007;18: 375–407. doi:10.1080/09548980701625173
62. Smith AC, Brown EN. Estimating a State-Space Model from Point Process Observations. Neural Comput. 2003;15: 965–991. doi:10.1162/089976603765202622
63. Vidne M, Ahmadian Y, Shlens J, Pillow JW, Kulkarni J, Litke AM, et al. Modeling the impact of common noise inputs on the network activity of retinal ganglion cells. J Comput Neurosci. 2012;33: 97–121. doi:10.1007/s10827-011-0376-2
64. Song D, Wang H, Tu CY, Marmarelis VZ, Hampson RE, Deadwyler SA, et al. Identification of sparse neural functional connectivity using penalized likelihood estimation and basis functions. J Comput Neurosci. 2013;35: 335–357. doi:10.1007/s10827-013-0455-7
65. Linderman SW, Adams RP, Pillow JW. Bayesian latent structure discovery from multi-neuron recordings. Neural Inf Process Syst. 2016; 11. Available: http://arxiv.org/abs/1610.08465
66. Robinson BS, Song D, Berger TW. Generalized Volterra kernel model identification of spike-timing-dependent plasticity from simulated spiking activity. Conf Proc . Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf. 2014;2014: 6585–6588. doi:10.1109/EMBC.2014.6945137
67. Stevenson I, Koerding K. Inferring spike-timing-dependent plasticity from spike train data. Adv Neural Inf Process Syst. 2011; 1–9. Available: http://machinelearning.wustl.edu/mlpapers/paper_files/NIPS2011_1399.pdf
68. Abbott LF, Regehr WG. Synaptic computation. Nature. 2004. doi:10.1038/nature03010
69. Zucker RS, Regehr WG. Short-term synaptic plasticity. Annu Rev Physiol. 2002;64: 355–405. doi:DOI 10.1146/annurev.physiol.64.092501.114547
70. Regehr WG. Short-term presynaptic plasticity. Cold Spring Harb Perspect Biol. 2012; doi:10.1101/cshperspect.a005702
71. Thomson AM, Lamy C. Functional maps of neocortical local circuitry. Front Neurosci. Frontiers; 2007;1: 19–42. doi:10.3389/neuro.01.1.1.002.2007
72. Dittman JS, Kreitzer AC, Regehr WG. Interplay between facilitation, depression, and residual calcium at three presynaptic terminals. J Neurosci. 2000;

doi:10.1523/JNEUROSCI.20-04-01374.2000

73. Wang Y, Markram H, Goodman PH, Berger TK, Ma J, Goldman-Rakic PS. Heterogeneity in the pyramidal network of the medial prefrontal cortex. Nat Neurosci. 2006;9: 534–542. doi:10.1038/nn1670

74. Fortune ES, Rose GJ. Short-term synaptic plasticity as a temporal filter. Trends Neurosci. 2001;24: 381–385. doi:10.1016/S0166-2236(00)01835-X

75. Cook DL, Schwindt PC, Grande LA, Spain WJ. Synaptic depression in the localization of sound. Nature. 2003;421: 66–70. doi:10.1038/nature01248

76. Abbott LF. Synaptic Depression and Cortical Gain Control. Science (80-). 1997;275: 221–224. doi:10.1126/science.275.5297.221

77. Drew PJ, Abbott LF. Extending the effects of spike-timing-dependent plasticity to behavioral timescales. Proc Natl Acad Sci. 2006;103: 8876–8881. doi:10.1073/pnas.0600676103

78. Usrey WM, Reppas JB, Reid RC. Paired-spike interactions and synaptic efficacy of retinal inputs to the thalamus. Nature. 1998; doi:10.1038/26487

79. Barthó P, Hirase H, Monconduit L, Zugaro M, Harris KD, Buzsáki G. Characterization of Neocortical Principal Cells and Interneurons by Network Interactions and Extracellular Features. J Neurophysiol. 2004;92: 600–608. doi:10.1152/jn.01170.2003

80. Csicsvari J, Hirase H, Czurko A, Buzsáki G. Reliability and state dependence of pyramidal cell-interneuron synapses in the hippocampus: An ensemble approach in the behaving rat. Neuron. 1998;21: 179–189. doi:10.1016/S0896-6273(00)80525-5

81. Fujisawa S, Amarasingham A, Harrison MT, Buzsáki G. Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. Nat Neurosci. 2008;11: 823–833. doi:10.1038/nn.2134

82. Carandini M, Horton JC, Sincich LC. Thalamic filtering of retinal spike trains by postsynaptic summation. J Vis. 2007;7: 20. doi:10.1167/7.14.20

83. Trussell LO, Zhang S, Ramant IM. Desensitization of AMPA receptors upon multiquantal neurotransmitter release. Neuron. 1993;10: 1185–1196. doi:10.1016/0896-6273(93)90066-Z

84. Henze DA, Buzsáki G. Action potential threshold of hippocampal pyramidal cells in vivo is increased by recent spiking activity. Neuroscience. 2001;105: 121–130.

doi:10.1016/S0306-4522(01)00167-1

85. Harris K, Csicsvari J, Hirase H, Dragoi G, Buzsáki G. Organization of cell assemblies in the hippocampus. Nature. 2003;424: 552–556. doi:10.1038/nature01765.1.

86. Ghanbari A, Malyshev A, Volgushev M, Stevenson IH. Estimating short-term synaptic plasticity from pre- and postsynaptic spiking. PLoS Comput Biol. 2017; doi:10.1371/journal.pcbi.1005738

87. Stoelzel CR, Bereshpolova Y, Gusev AG, Swadlow HA. The Impact of an LGNd Impulse on the Awake Visual Cortex: Synaptic Dynamics and the Sustained/Transient Distinction. J Neurosci. 2008; doi:10.1523/JNEUROSCI.4726-07.2008

88. Stoelzel CR, Bereshpolova Y, Swadlow HA. Stability of Thalamocortical Synaptic Transmission across Awake Brain States. J Neurosci. 2009; doi:10.1523/JNEUROSCI.5983-08.2009

89. Thomson AM, West DC, Wang Y, Bannister AP. Synaptic Connections and Small Circuits Involving Excitatory and Inhibitory Neurons in Layers 2-5 of Adult Rat and Cat Neocortex: Triple Intracellular Recordings and Biocytin Labelling In Vitro. Cereb Cortex. Oxford University Press; 2002;12: 936–953. doi:10.1093/cercor/12.9.936

90. Yang H, Xu-Friedman MA. Impact of Synaptic Depression on Spike Timing at the Endbulb of Held. J Neurophysiol. 2009; doi:10.1152/jn.00072.2009

91. Yang H, Xu-Friedman MA. Relative Roles of Different Mechanisms of Depression at the Mouse Endbulb of Held. J Neurophysiol. 2008; doi:10.1152/jn.01293.2007

92. Stevenson IH, London BM, Oby ER, Sachs NA, Reimer J, Englitz B, et al. Functional Connectivity and Tuning Curves in Populations of Simultaneously Recorded Neurons. PLoS Comput Biol. Public Library of Science; 2012;8: e1002775. doi:10.1371/journal.pcbi.1002775

93. Swadlow HA, Gusev AG. Receptive-field construction in cortical inhibitory interneurons. Nature Neuroscience. 2002. doi:10.1038/nn847

94. Kawaguchi Y. Distinct firing patterns of neuronal subtypes in cortical synchronized activities. J Neurosci. 2001; doi:21/18/7261 [pii]

95. Keine C, Rübsamen R, Englitz B. Signal integration at spherical bushy cells enhances representation of temporal structure but limits its range. Elife. 2017;6. doi:10.7554/eLife.29639

96. Keine C, Rubsamen R, Englitz B. Inhibition in the auditory brainstem enhances signal representation and regulates gain in complex acoustic environments. Elife. 2016;5. doi:10.7554/eLife.19295

97. Mora Lopez C, Putzeys J, Raducanu BC, Ballini M, Wang S, Andrei A, et al. A Neural Probe with Up to 966 Electrodes and Up to 384 Configurable Channels in 0.13 μm SOI CMOS. IEEE Trans Biomed Circuits Syst. 2017;11: 510–522. doi:10.1109/TBCAS.2016.2646901

98. Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, et al. Fully integrated silicon probes for high-density recording of neural activity. Nature. 2017;551: 232–236. doi:10.1038/nature24636

99. Pachitariu M, Steinmetz N, Kadir S, Carandini M, Harris KD. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels [Internet]. bioRxiv. 2016. doi:10.1101/061481

100. Moore GP, Segundo JP, Perkel DH, Levitan H. Statistical Signs of Synaptic Interaction in Neurons. Biophys J. 1970;10: 876–900. doi:10.1016/S0006-3495(70)86341-X

101. Dale H. Pharmacology and Nerve-endings. J R Soc Med. 1935;28: 319–332. doi:10.1177/003591573502800330

102. Tsodyks M V., Markram H. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. Proc Natl Acad Sci. 1997;94: 719–723. doi:10.1073/pnas.94.2.719

103. Hatsopoulos NG, Xu Q, Amit Y. Encoding of Movement Fragments in the Motor Cortex. J Neurosci. 2007;27: 5105–5114. doi:10.1523/JNEUROSCI.3570-06.2007

104. Huang C, Resnik A, Celikel T, Englitz B. Adaptive Spike Threshold Enables Robust and Temporally Precise Neuronal Encoding. PLoS Comput Biol. 2016; doi:10.1371/journal.pcbi.1004984

105. Gil Z, Connors BW, Amitai Y. Differential regulation of neocortical synapses by neuromodulators and activity. Neuron. 1997; doi:10.1016/S0896-6273(00)80380-3

106. Wang Y, Manis PB. Synaptic Transmission at the Cochlear Nucleus Endbulb Synapse During Age-Related Hearing Loss in Mice. J Neurophysiol. 2005; doi:10.1152/jn.00374.2005

107. Wang Y, Manis PB. Short-term synaptic depression and recovery at the mature mammalian

endbulb of Held synapse in mice. J Neurophysiol. 2008; doi:10.1152/jn.90715.2008

108. Paz JT, Chavez M, Saillet S, Deniau J-M, Charpier S. Activity of Ventral Medial Thalamic Neurons during Absence Seizures and Modulation of Cortical Paroxysms by the Nigrothalamic Pathway. J Neurosci. 2007; doi:10.1523/JNEUROSCI.4677-06.2007
109. Moore AK, Wehr M. Parvalbumin-Expressing Inhibitory Interneurons in Auditory Cortex Are Well-Tuned for Frequency. J Neurosci. 2013;33: 13713–13723. doi:10.1523/JNEUROSCI.0663-13.2013
110. Perrenoud Q, Rossier J, Geoffroy H, Vitalis T, Gallopin T. Diversity of gabaergic interneurons in layer VIa and VIb of mouse barrel cortex. Cereb Cortex. 2013; doi:10.1093/cercor/bhs032
111. London M, Schreibman A, Haä;usser M, Larkum ME, Segev I. The information efficacy of a synapse. Nat Neurosci. 2002;5: 332–340. doi:10.1038/nn826
112. Pinault D. The juxtacellular recording-labeling technique. Neuromethods. 2011. doi:10.1007/978-1-60327-202-5_3
113. Weiler N, Wood L, Yu J, Solla SA, Shepherd GMG. Top-down laminar organization of the excitatory network in motor cortex. Nat Neurosci. 2008; doi:10.1038/nn2049
114. Ladenbauer J, McKenzie S, English DF, Hagens O, Ostojic S. Inferring and validating mechanistic models of neural microcircuits based on spike-train data. bioRxiv Prepr. 2018; doi:10.1101/261016
115. Robinson BS, Berger TW, Song D. Identification of stable spike-timing-dependent plasticity from spiking activity with generalized multilinear modeling. Neural Comput. 2016; doi:10.1162/NECO_a_00883
116. Song D, Robinson BS, Berger TW. Identification of Short-Term and Long-Term Functional Synaptic Plasticity From Spiking Activities. Adaptive Learning Methods for Nonlinear System Modeling. Elsevier; 2018. pp. 289–312. doi:10.1016/B978-0-12-812976-0.00017-8
117. Amidi Y, Nazari B, Sadri S, … UE-2018 40th A, 2018 U. Parameter Estimation in Synaptic Coupling Model Using a Point Process Modeling Framework*. ieeexplore.ieee.org. 2018; doi:10.1109/EMBC.2018.8512815
118. Bayat Mokhtari E, Lawrence JJ, Stone EF. Data Driven Models of Short-Term Synaptic Plasticity. Front Comput Neurosci. 2018; doi:10.3389/fncom.2018.00032

119. Karmarkar UR, Buonomano D V. Timing in the Absence of Clocks: Encoding Time in Neural Network States. Neuron. 2007; doi:10.1016/j.neuron.2007.01.006