

Novel Tools and Methods

Quantifying Repetitive Transmission at Chemical Synapses: A Generative-Model Approach^{1,2,3}

Alessandro Barri,¹ Yun Wang,² David Hansel,^{3,4} and Gianluigi Mongillo^{3,4}DOI:<http://dx.doi.org/10.1523/ENEURO.0113-15.2016>

¹Unité d'Imagerie Dynamique du Neurone, Institut Pasteur, Paris, France, ²Caritas St. Elizabeth's Center, Tufts University, Boston, MA, USA, ³Centre National de la Recherche Scientifique, UMR 8119, Paris, France, ⁴Cerebral Dynamics, Plasticity and Learning, Centre de Neurophysique, Physiologie et Pathologie, Université Descartes, Paris, France

Abstract

The dependence of the synaptic responses on the history of activation and their large variability are both distinctive features of repetitive transmission at chemical synapses. Quantitative investigations have mostly focused on trial-averaged responses to characterize dynamic aspects of the transmission—thus disregarding variability—or on the fluctuations of the responses in steady conditions to characterize variability—thus disregarding dynamics. We present a statistically principled framework to quantify the dynamics of the probability distribution of synaptic responses under arbitrary patterns of activation. This is achieved by constructing a generative model of repetitive transmission, which includes an explicit description of the sources of stochasticity present in the process. The underlying parameters are then selected via an expectation-maximization algorithm that is exact for a large class of models of synaptic transmission, so as to maximize the likelihood of the observed responses. The method exploits the information contained in the correlation between responses to produce highly accurate estimates of both quantal and dynamic parameters from the same recordings. The method also provides important conceptual and technical advances over existing state-of-the-art techniques. In particular, the repetition of the same stimulation in identical conditions becomes unnecessary. This paves the way to the design of optimal protocols to estimate synaptic parameters, to the quantitative comparison of synaptic models over benchmark datasets, and, most importantly, to the study of repetitive transmission under physiologically relevant patterns of synaptic activation.

Key words: expectation-maximization; generative modeling; quantal analysis; repetitive transmission; short-term plasticity

Significance Statement

Transmission at chemical synapses is transiently adjusted on a spike-by-spike basis, which has been proposed to enhance information processing in neuronal networks. So far, however, the dynamic properties of transmission have been characterized only for physiologically unrealistic patterns of activation. This is because the current methods used to estimate the parameters describing repetitive transmission are unable to deal with the fluctuations of the responses. These must either be averaged out or estimated directly from the data, which requires a large number of repetitions of the same stimulation, severely constraining experimental protocols. We developed a novel method that allows one to estimate the parameters from a single, arbitrary pattern of activation. The method lays the groundwork for the characterization of transmission with *in vivo*-like patterns of activation.

Introduction

A distinctive feature of chemical transmission is the rapid and transient modification of the postsynaptic response as a result of repetitive presynaptic activation (Zucker and Regehr, 2002; Fioravante and Regehr, 2011). The ability of chemical synapses to quickly adjust their transmission properties in an activity-dependent way has been suggested to significantly enhance information processing in neuronal networks. Important computations are thought to rely, fully or partly, on “synaptic computations” (Abbott and Regehr, 2004; Wu et al., 2013). A nonexhaustive list includes the following: rhythm generation (Senn et al., 1996; Tsodyks et al., 2000), gain control (Abbott et al., 1997; Rothman et al., 2009), temporal filtering (Fortune and Rose, 2001), temporary memory maintenance (Hempel et al., 2000; Barak and Tsodyks, 2007; Mongillo et al., 2008), and the source of nonlinearity in the balanced regime (Mongillo et al., 2012; Hansel and Mato, 2013).

Quantitative investigation of repetitive synaptic transmission largely relies on phenomenological descriptions (Bertram et al., 1996; Tsodyks and Markram, 1997; Varela et al., 1997; Markram et al., 1998; Dittman et al., 2000). In their providing compact (i.e., with few parameters), low-dimensional descriptions, phenomenological models have been instrumental in effectively classifying patterns of transmission at different synapses (Gupta et al., 2000; Blackman et al., 2013), in uncovering their underlying mechanisms (Dittman et al., 2000; Saviane and Silver, 2006; Hallermann et al., 2010b), and in exploring theoretically their functional consequences. Phenomenological models, however, only describe the average responses or, where the model is stochastic, it is the average model responses that are fitted to the trial-averaged experimental responses. In either case, the trial-to-trial variability and the within-trial correlation of the responses are neglected (but, see Costa et al., 2013). The variability of the synaptic responses is, indeed, another distinctive feature of chemical transmission. This variability is understood and routinely quantified in terms of the quantal model of synaptic release (Quastel, 1997; Stevens, 2003). Methods

to estimate quantal parameters are, however, tailored for steady-state conditions, and their extension to dynamic conditions has proven difficult (Scheuss et al., 2002; Saviane and Silver, 2006; Loebel et al., 2009; Hallermann et al., 2010a,b).

The trial-averaging procedure required to fit models to data destroys the large amount of information contained in the correlation between consecutive responses as well as in their fluctuations. The accuracy of the parameters estimate, achievable by least-squares fitting, is thus seriously limited and steadily declines with increasing the complexity of the model (i.e., with the number of parameters to be fitted). Trial averaging also severely constrains experimental protocols. The need to have a suitable number of repetitions in identical conditions leads, essentially, to protocols consisting of short, regular presynaptic trains at relatively high rates, followed by quite long interstimulation intervals (compared with stimulation periods). The repetition of identical trains (regular or not) allows one to extract very little information about the underlying synaptic dynamics. Moreover, the parameters are estimated with patterns of synaptic activation that are arguably very far from physiological patterns, raising the question of how good a description are the current models and/or parameters for repetitive synaptic transmission in *in vivo*-like conditions (Dobrunz and Stevens, 1999; Kandaswamy et al., 2010).

Here, we provide a new methodology that integrates information about the stochasticity of the synaptic responses in the estimation of the parameters describing the dynamic properties of synaptic transmission. We introduce a novel class of generative models that allows one to compactly describe the statistical dependencies among consecutive synaptic responses as well as their variability. We show that both inference and learning are tractable in this class of generative models. In particular, we develop an exact expectation-maximization (EM) algorithm that allows one to estimate the parameters for arbitrary patterns of presynaptic stimulation. We demonstrate two main advantages of our approach over conventional techniques. First, we simultaneously estimate both quantal and dynamic parameters from the same experimental recordings. Second, and most importantly, since the estimation procedure does not rely on trial-averaged quantities, parameters can be estimated from single traces. It is thus possible to devise alternative stimulation protocols and analyze their impact on parameter estimation by the use of theoretical tools.

Materials and Methods

Electrophysiological recordings

We refer the reader to Wang et al. (2006) for a detailed description. Briefly, acute slices were cut from the medial prefrontal cortex of young ferrets (1.5–3 months old), and whole-cell patch-clamp recordings were made from synaptically connected layer 5 pyramidal neurons. Synaptic transmission was probed by eliciting in the presynaptic cell regular trains of five or eight spikes at varying frequencies, followed by a recovery spike. Postsynaptic responses were recorded in current-clamp mode. The

Received September 24, 2015; accepted April 2, 2016; First published April 29, 2016.

¹The authors declare no competing financial interests.

²Author Contributions: A.B., D.H., and G.M. designed research; A.B., D.H., and G.M. performed research; A.B., Y.W., D.H., and G.M. analyzed data; A.B., Y.W., D.H., and G.M. wrote the paper.

³This research was conducted within the scope of the France-Israel Laboratory of Neuroscience (FILN-LEA), and supported by the France-Israel High Council for Science and Technology, the French National Research Agency (ANR) and the joint CNRS-JSPS research project.

Acknowledgements: We are indebted to Carl van Vreeswijk for many insightful discussions and to David DiGregorio and German Mato for many useful comments on a previous version of the manuscript.

Correspondence should be addressed to Gianluigi Mongillo, CNRS and Paris Descartes University, 45, rue des Saints-Pères, Paris 75006, France. E-mail: gianluigi.mongillo@parisdescartes.fr.

DOI: <http://dx.doi.org/10.1523/ENEURO.0113-15.2016>

Copyright © 2016 Barri et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

interspike interval of the train T ranged between 14.3 and 200 ms (5–70 Hz). The interval for the recovery spike was correspondingly determined as $T_{\text{rec}} = T + 500$ ms. This stimulation protocol was repeated between 20 and 40 times.

Most of the connections were probed with a single frequency of stimulation (20 Hz, i.e., $T = 50$ ms). For the connections probed with several frequencies, we retained only one for further analysis. The retained frequency was chosen so as to have the largest possible number of connections for each of the frequencies used in the experiment.

Preprocessing

Postsynaptic voltage traces that exhibited drift or an abrupt change in the baseline during recording were excluded from further analysis. The remaining voltage traces were smoothed by using a rectangular window 2 ms in size. From the smoothed traces, we extracted the single peak responses using the method described by [Richardson and Silberberg \(2008\)](#). Briefly, from the trial-averaged trace we estimated the membrane time constant τ_m by fitting an exponential decay to the falling edge of the recovery response and, where possible, also to the falling edge of the first response, and averaged over these. Each voltage trace $V(t)$ was then deconvolved using the following:

$$RI(t) = \tau_m \cdot \frac{dV(t)}{dt} + V(t). \quad (1)$$

From the $RI(t)$ trace, which typically featured clearly separated peaks, we cropped 12 ms windows centered around the nominal time of the presynaptic stimulation. These “crops” were then reconvolved, yielding fully separated excitatory postsynaptic potentials, from which we finally extracted the peak responses. From the “crops” we also estimated the baseline noise σ_r^2 by computing the variance of the voltage trace over a 1 ms time window before the onset of the response. The baseline noise was estimated separately for each connection.

The stochastic Tsodyks-Markram model

The stochastic Tsodyks-Markram (TM) model ([Fuhmann et al., 2002](#)) describes the synapse as a collection of N identical and independent release sites. Each site can be in one of the two following states: competent (i.e., occupied by one vesicle), or noncompetent (i.e., no vesicle is docked to the site). Upon spike, a competent site can probabilistically release the vesicle, and become noncompetent. The probability of release $u(t)$ depends on the history of synaptic activation according to the following:

$$\dot{u} = \frac{U-u}{\tau_F} + U \cdot (1-u) \sum_k \delta(t-t_k), \quad (2)$$

where U is the initial release probability, τ_F is the time constant of facilitation, and the sum over k is over all presynaptic spike times, t_k . In between spikes, vesicles dock probabilistically to noncompetent sites. The proba-

bility of docking within a time interval Δ since the site first became noncompetent, $l(\Delta)$, is given by the following:

$$l(\Delta) = 1 - e^{-\Delta/\tau_D}, \quad (3)$$

where τ_D is the average time a vesicle takes to dock to a noncompetent site. The response R to one vesicle is a random variable with mean q (quantal size) and variance σ_q^2 (quantal variability). To account for the right-skewness of unitary quantal responses ([Bekkers et al., 1990](#); [Bhumbra and Beato, 2013](#)), we use an inverse Gaussian for the probability distribution function of R , as follows:

$$P(R) = \frac{q^{3/2}}{\sqrt{2\pi\sigma_q^2 R^3}} \exp\left[-\frac{q(R-q)^2}{2\sigma_q^2 R}\right]. \quad (4)$$

The response to more than one vesicle is the linear sum of single quantal responses (see [Eq. 11](#)). The average postsynaptic responses \bar{R}_k elicited by presynaptic spikes at times t_k are given by the following:

$$\bar{R}_k = A \cdot u_k \cdot x_k, \quad (5)$$

where $A \equiv N \cdot q$ is the so-called absolute synaptic efficacy, u_k is the probability of release immediately before spike at time t_k , and x_k is the probability that a site is release competent immediately before the spike at time t_k . Both u_k and x_k can be recursively evaluated as follows:

$$u_{k+1} = U + u_k \cdot (1 - U) \cdot \exp\left(-\frac{\Delta_k}{\tau_F}\right) \quad (6)$$

$$x_{k+1} = 1 - [1 - (1 - u_k) \cdot x_k] \cdot \exp\left(-\frac{\Delta_k}{\tau_D}\right), \quad (7)$$

where $\Delta_k \equiv t_{k+1} - t_k$ is the k th interspike interval, $u_1 = U$ and $x_1 = 1$.

Likelihood of a response sequence

The likelihood of observing a sequence of postsynaptic responses $R_{1 \rightarrow M} \equiv \{R_1, \dots, R_M\}$ elicited by a train of presynaptic spikes occurring at times $t_{1 \rightarrow M} \equiv \{t_1, \dots, t_M\}$ is given by the following:

$$L(\theta | R_{1 \rightarrow M}) \equiv P(R_{1 \rightarrow M} | \theta) = \sum_{S_{1 \rightarrow M}^-, S_{1 \rightarrow M}^+} P(R_{1 \rightarrow M}, S_{1 \rightarrow M}^-, S_{1 \rightarrow M}^+ | \theta), \quad (8)$$

where $S_{1 \rightarrow M}^-$ and $S_{1 \rightarrow M}^+$ denote the sequences of a number of release-competent sites immediately before and after each spike, respectively. The model parameters are denoted by θ . For concreteness, in the case of the stochastic TM model they are as follows: the number of release sites N , the quantal size q , the quantal noise σ_q^2 , the initial release probability U , and the time constants for docking and facilitation τ_D and τ_F , respectively. To simplify the notation, we omit hereafter the dependence on the parameters θ assumed to be constant in what follows. The joint probability of $R_{1 \rightarrow M}$, $S_{1 \rightarrow M}^-$ and $S_{1 \rightarrow M}^+$ can be written as follows:

$$P(R_{1 \rightarrow M}, S_{1 \rightarrow M}^-, S_{1 \rightarrow M}^+) = P(S_{\bar{1}}^-) \prod_{k=1}^M P(S_k^+ | S_k^-) P(R_k | S_k^+, S_k^-) \prod_{k=1}^{M-1} P(S_{k+1}^- | S_k^+), \quad (9)$$

where $P(S_{\bar{1}}^-)$ is the probability that $S_{\bar{1}}^-$ sites are competent at the beginning of the spike train. For the stochastic TM model, all sites are release competent in the absence of stimulation. We have $P(S_{\bar{1}}^-) = 1$ if $S_{\bar{1}}^- = N$ and $P(S_{\bar{1}}^-) = 0$ otherwise. The term $P(S_k^+ | S_k^-)$ describes the release process (i.e., it is the probability that the number of release-competent sites changes from S_k^- to S_k^+ upon the k th spike). Before the k th spike, there are S_k^- release-competent sites that can independently release with probability u_k . The number of release-competent sites cannot increase upon spike. Thus:

$$P(S_k^+ | S_k^-) = \begin{cases} \binom{S_k^-}{S_k^- - S_k^+} \cdot u_k^{(S_k^- - S_k^+)} \cdot (1 - u_k)^{S_k^+} & \text{if } S_k^+ \leq S_k^- \\ 0 & \text{otherwise} \end{cases}. \quad (10)$$

The term $P(R_k | S_k^+, S_k^-)$ describes the postsynaptic responses (i.e., it is the probability of observing a response between R_k and $R_k + dR$ when the number of release-competent sites changes from S_k^- to S_k^+ upon the k th spike). Note that $S_k^- - S_k^+$ is the number of vesicles released. Because of linear summation, R_k is also distributed according to an inverse Gaussian distribution with mean $(S_k^- - S_k^+) \cdot q$ and variance $(S_k^- - S_k^+) \cdot \sigma_q^2$, as follows:

$$P(R_k | S_k^+, S_k^-) = \frac{q^3 \cdot (S_k^- - S_k^+)}{\sqrt{2\pi\sigma_q^2 R_k^3}} \times \exp\left\{-\frac{q \cdot [R_k - q \cdot (S_k^- - S_k^+)]^2}{2\sigma_q^2 R_k}\right\}. \quad (11)$$

Finally, the term $P(S_{k+1}^- | S_k^+)$ describes the docking process, i.e., it is the probability that the number of release-competent sites changes from S_k^+ to S_{k+1}^- during the k th interspike interval Δ_k . After the k th spike, there are $N - S_k^+$ noncompetent sites that can independently become release competent within the time interval Δ_k with the probability $l_k \equiv l(\Delta_k)$ (see Eq. 3). The number of release-competent sites cannot decrease in between spikes. Thus:

$$P(S_{k+1}^- | S_k^+) = \begin{cases} \binom{N - S_k^+}{S_{k+1}^- - S_k^+} \cdot l_k^{(S_{k+1}^- - S_k^+)} \cdot (1 - l_k)^{(N - S_{k+1}^-)} & \text{if } S_{k+1}^- \geq S_k^+ \\ 0 & \text{otherwise} \end{cases}. \quad (12)$$

The sum over all possible realizations of $S_{1 \rightarrow M}^-$ and $S_{1 \rightarrow M}^+$ in Equation 8 can be efficiently performed as described in the Forward-backward formalism section.

Expectation-Maximization

The fact that the number of release-competent sites is not directly observable makes the direct maximization of the likelihood function in Equation 8 impractical. The EM algorithm allows for maximum-likelihood estimation in the presence of hidden variables (Dempster et al., 1977; Dayan and Abbott, 2001). EM iteratively improves upon an initial guess of the parameters by maximizing the so-called auxiliary function with respect to θ . The auxiliary function is defined as follows:

$$Q(\theta, \theta^{old}) = \sum_{S_{1 \rightarrow M}^-} \sum_{S_{1 \rightarrow M}^+} P(S_{1 \rightarrow M}^-, S_{1 \rightarrow M}^+ | R_{1 \rightarrow M}, \theta^{old}) \times \log[P(R_{1 \rightarrow M}, S_{1 \rightarrow M}^-, S_{1 \rightarrow M}^+ | \theta)], \quad (13)$$

where θ^{old} is the initial guess for the parameters, and the sum is over all possible sequences $S_{1 \rightarrow M}^-$ and $S_{1 \rightarrow M}^+$. The EM algorithm comprises two steps. The so-called E step corresponds to the evaluation of the auxiliary function, that is, the computation of the expectation of $\log[P(R_{1 \rightarrow M}, S_{1 \rightarrow M}^-, S_{1 \rightarrow M}^+ | \theta)]$ given the observed responses and the current estimate. The so-called M step corresponds to maximizing the auxiliary function with respect to θ , as follows:

$$\theta^{new} = \arg \max_{\theta} Q(\theta, \theta^{old}). \quad (14)$$

Iteration of the E and M steps guarantees improvement of the initial guess until eventually a fixed point is reached (i.e., $\theta^{new} = \theta^{old}$), which corresponds to a (local) maximum of the likelihood function.

Using Equations 9–12 in Equation 13, taking the derivatives with respect to the continuous parameters, and setting them to 0, we obtain the following re-estimation formulas:

$$\sum_{k=1}^M \langle R_k - q^{new} \cdot (S_k^- - S_k^+) \rangle = 0 \quad [q^{new}]. \quad (15)$$

$$\sum_{k=1}^M \left\langle 1 - \frac{q^{new}}{(\sigma_q^{new})^2 R_k} \cdot (R_k - q^{new} \cdot (S_k^- - S_k^+))^2 \right\rangle = 0 \quad [\sigma_q^{new}] \quad (16)$$

$$\sum_{k=1}^{M-1} \frac{\partial l_k}{\partial \tau_D} \left[\left\langle \frac{S_{k+1}^-}{l_k(1 - l_k)} \right\rangle - \left\langle \frac{S_k^+}{l_k} \right\rangle - \frac{N}{1 - l_k} \right] = 0 \quad [\tau_D^{new}] \quad (17)$$

$$\sum_{k=1}^M \frac{\partial u_k}{\partial U} \left[\left\langle \frac{S_k^-}{u_k} \right\rangle - \left\langle \frac{S_k^+}{u_k(1 - u_k)} \right\rangle \right] = 0 \quad [U^{new}] \quad (18)$$

$$\sum_{k=1}^M \frac{\partial u_k}{\partial \tau_F} \left[\left\langle \frac{S_k^-}{u_k} \right\rangle - \left\langle \frac{S_k^+}{u_k(1 - u_k)} \right\rangle \right] = 0. \quad [\tau_F^{new}]. \quad (19)$$

Note that u_k and l_k depend on the new estimate U^{new} , τ_F^{new} , and τ_D^{new} . The angular brackets denote average over

the distribution $P(S_{1 \rightarrow M}^-, S_{1 \rightarrow M}^+ | R_{1 \rightarrow M}, \theta^{old})$ (see Eq. 13). This average has the following convenient property:

$$\begin{aligned} \langle g(S_k^-, S_k^+) \rangle &= \sum_{S_{1 \rightarrow M}^-} \sum_{S_{1 \rightarrow M}^+} g(S_k^-, S_k^+) \cdot P(S_{1 \rightarrow M}^-, S_{1 \rightarrow M}^+ | R_{1 \rightarrow M}, \theta^{old}) \\ &= \sum_{S_k^-, S_k^+} g(S_k^-, S_k^+) \cdot P(S_k^-, S_k^+ | R_{1 \rightarrow M}, \theta^{old}). \end{aligned} \quad (20)$$

which holds for any function, g . The conditions for U^{new} and τ_F^{new} involve u_k and its derivatives, which in turn depend on U^{new} and τ_F^{new} . These conditions thus have to be evaluated simultaneously.

In the presence of Gaussian baseline noise (independent of the number of released vesicles), Equation 11 becomes the following:

$$\begin{aligned} &P(R_k | S_k^+, S_k^-) \\ &= \int_0^{+\infty} \frac{dy}{\sigma_n \sqrt{2\pi}} P(y | S_k^+, S_k^-) \cdot \exp\left[-\frac{(R_k - y)^2}{2\sigma_n^2}\right]. \end{aligned} \quad (21)$$

where σ_n^2 is the estimated variance of the baseline noise (see the Preprocessing section). When the above equation is used to derive the EM re-estimation formulas, the equations for q^{new} and σ_q^{new} read as follows:

$$\sum_{k=1}^M \left\langle \frac{\int_0^{+\infty} y \cdot l(y) dy}{\int_0^{+\infty} l(y) dy} - q^{new} \cdot (S_k^- - S_k^+) \right\rangle = 0 \quad [q^{new}] \quad (22)$$

$$\sum_{k=1}^M \left\langle \frac{\int_0^{+\infty} (y - q^{new} \cdot (S_k^- - S_k^+))^2 \cdot y^{-1} \cdot l(y) dy}{\int_0^{+\infty} l(y) dy} - \frac{(\sigma_q^{new})^2}{q^{new}} \right\rangle = 0 \quad [\sigma_q^{new}]. \quad (23)$$

with $l(y)$ given by:

$$l(y) = y^{-\frac{3}{2}} \cdot \exp\left\{-\frac{q \cdot [y - q \cdot (S_k^- - S_k^+)]^2}{2\sigma_q^2 y} - \frac{(R_k - y)^2}{2\sigma_n^2}\right\}. \quad (24)$$

Re-estimation formulas have to be evaluated numerically. The EM algorithm was implemented in C++. For multidimensional root finding, we used the Brent-Dekker algorithm and a derivative-free version of Powell's hybrid algorithm, as provided by the GNU Scientific Library (Galassi et al., 2009). The code is available from the authors upon request.

Forward-backward formalism

To efficiently compute the averages over $P(S_{1 \rightarrow M}^-, S_{1 \rightarrow M}^+ | R_{1 \rightarrow M}, \theta^{old})$, we developed a forward-backward scheme (Rabiner, 1989). We define two forward variables as follows:

$$\alpha_k^-(S) = P(S_k^- = S, R_{1 \rightarrow k-1}) \quad (25)$$

$$\alpha_k^+(S) = P(S_k^+ = S, R_{1 \rightarrow k}), \quad (26)$$

where we have dropped the dependence of the parameters to simplify the notation. These variables can be evaluated recursively as follows:

$$\alpha_1^-(S) = P(S_1^- = S) \quad (27)$$

$$\alpha_k^+(S) = \sum_{S_k^-=0}^N \alpha_k^-(S) P(S_k^+ = S | S_k^-) P(R_k | S_k^+ = S, S_k^-) \quad (28)$$

$$\alpha_{k+1}^-(S) = \sum_{S_k^+=0}^N \alpha_k^+(S) P(S_{k+1}^- = S | S_k^+). \quad (29)$$

Note that $\alpha_M^+(S) = P(S_M^+ = S, R_{1 \rightarrow M})$ and thus $\sum_{S=0}^N \alpha_M^+(S) = P(R_{1 \rightarrow M})$. Similarly, we can define two backward variables as follows:

$$\beta_k^-(S) = P(R_{k \rightarrow M} | S_k^- = S) \quad (30)$$

$$\beta_k^+(S) = P(R_{k+1 \rightarrow M} | S_k^+ = S), \quad (31)$$

which can be evaluated recursively as follows:

$$\beta_1^+(S) = 1 \quad (32)$$

$$\beta_k^-(S) = \sum_{S_k^+=0}^N \beta_k^+(S) P(S_k^+ | S_k^- = S) P(R_k | S_k^+, S_k^- = S) \quad (33)$$

$$\beta_{k-1}^+(S) = \sum_{S_k^-=0}^N \beta_k^-(S) P(S_k^- | S_{k-1}^+ = S), \quad (34)$$

and $\sum_{S=0}^N \beta_1^-(S) P(S_1^- = S) = P(R_{1 \rightarrow M})$. From this, it follows that the conditional distribution appearing in Equation 20 can be easily computed from the following:

$$\begin{aligned} &P(S_k^-, S_k^+ | R_{1 \rightarrow M}, \theta) \\ &= \frac{\beta_k^+(S_k^+) \cdot P(R_k | S_k^-, S_k^+) \cdot P(S_k^+ | S_k^-) \cdot \alpha_k^-(S_k^-)}{P(R_{1 \rightarrow M})}. \end{aligned} \quad (35)$$

Fisher information matrix

To quantify the amount of information about a specific parameter that a stimulation protocol allows one to extract, we can compute the Fisher Information Matrix (FIM) of the generative model

$$\begin{aligned} I(\theta)_{j,k} &= E\left[\left(\frac{\partial}{\partial \theta_j} \log\{P(R_{1 \rightarrow M} | t_{1 \rightarrow M}, \theta)\}\right) \right. \\ &\quad \left. \times \left(\frac{\partial}{\partial \theta_k} \log\{P(R_{1 \rightarrow M} | t_{1 \rightarrow M}, \theta)\}\right) \middle| \theta \right], \end{aligned} \quad (36)$$

where we have introduced explicitly the dependence on $t_{1 \rightarrow M}$, $E[\cdot]$ that denotes the expectation value over the responses sequences, and j and k denote (continuous) model parameters. The diagonal elements of the inverse FIM are lower bounds on the variances of the corresponding parameters estimate. Hence, a lower bound on the relative error ϵ of the estimate of θ_j can be written as follows:

$$\epsilon(\theta) \geq \frac{\sqrt{[\mathbf{I}(\theta)^{-1}]_{jj}}}{\theta_j}, \quad (37)$$

which is just a normalized version of the Cramér-Rao bound (Dayan and Abbott, 2001).

The calculation of the derivatives in Equation 36 can be performed efficiently using the forward variables $\alpha_k^\pm(S)$. We can write the following:

$$\frac{\partial}{\partial \theta_j} \log[P(R_{1 \rightarrow M} | t_{1 \rightarrow M}, \theta)] = \frac{1}{P(R_{1 \rightarrow M} | t_{1 \rightarrow M}, \theta)} \cdot \sum_{S=0}^N \frac{\partial \alpha_M^+(S)}{\partial \theta_j}. \quad (38)$$

Since $\alpha_k^+(S)$ depends on $\alpha_k^-(S)$ and vice versa, we obtain recursive formulas for their respective derivatives, as follows:

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \alpha_k^+(S) &= \sum_{S_k^-=0}^N \alpha_k^-(S_k^-) \cdot P(S_k^+ = S | S_k^-) \cdot \\ &\left\{ \frac{\partial P(R_k | S_k^-, S_k^+ = S)}{\partial \theta_j} \right\} + \sum_{S_k^-=0}^N \alpha_k^-(S_k^-) \cdot \\ &\left\{ \frac{\partial P(S_k^+ = S | S_k^-)}{\partial \theta_j} \right\} \cdot P(R_k | S_k^-, S_k^+ = S) + \sum_{S_k^-=0}^N \\ &\left\{ \frac{\partial \alpha_k^-(S_k^-)}{\partial \theta_j} \right\} \cdot P(S_k^+ = S | S_k^-) \cdot P(R_k | S_k^-, S_k^+ = S), \quad (39) \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \alpha_k^-(S) &= \sum_{S_{k-1}^+=0}^N \alpha_{k-1}^+(S_{k-1}^+) \cdot \left\{ \frac{\partial P(S_k^- = S | S_{k-1}^+)}{\partial \theta_j} \right\} \\ &+ \sum_{S_{k-1}^+=0}^N \left\{ \frac{\partial \alpha_{k-1}^+(S_{k-1}^+)}{\partial \theta_j} \right\} \cdot P(S_k^- = S | S_{k-1}^+). \quad (40) \end{aligned}$$

Least-squares fitting

To fit the average model response $\bar{R}_{1 \rightarrow M}$ to the average experimental responses $\langle R \rangle_{1 \rightarrow M}$, we used a standard least-squares procedure (see, e.g., Tsodyks and Markram, 1997; Markram et al., 1998), as follows:

$$\hat{\theta}_{LS} = \arg \min_{\theta_{LS}} \sum_{i=1}^M \frac{(\langle R \rangle_i - \bar{R}_i)^2}{s_i^2}. \quad (41)$$

where the \bar{R}_i values depends on the parameters $\theta_{LS} = \{A, U, \tau_D, \tau_F\}$ (see Eqs. 5–7), and s_i^2 denotes the variance of the i th response.

Least-squares condition number

The least-squares fitting is a mapping from the average experimental responses $\langle R \rangle_{1 \rightarrow M}$ to the four parameters $\theta_{LS} = \{A, U, \tau_D, \tau_F\}$, as follows:

$$\theta_{LS} = F(\langle R \rangle_{1 \rightarrow M}). \quad (42)$$

The condition number (Trefethen and Bau, 1997) of this mapping is given by the following:

$$c = \frac{\|D\| \cdot \|\bar{R}_{1 \rightarrow M}\|}{\|\theta_{LS}\|}, \quad (43)$$

where $\|\cdot\|$ denotes the 2-norm and D is a $4 \times M$ matrix whose elements are as follows:

$$D_{ij} = \frac{\partial F_i}{\partial \langle R \rangle_j} = \frac{\partial \theta_{LS}^i}{\partial \langle R \rangle_j}. \quad (44)$$

The D_{ij} can be straightforwardly evaluated numerically. The condition number measures the sensitivity of the estimates to small changes in the average responses. If $c \gg 1$, small perturbations in the average responses will tend to cause disproportionately large changes in the estimates of the parameters. In that case, the estimation procedure is ill posed.

Results

Generative model description of repetitive synaptic transmission: overview

We provide here a compact presentation of our method omitting unnecessary technical details. The full presentation can be found in Materials and Methods.

Synaptic transmission is, at least in the experimental conditions in which it is routinely probed, stochastic so that recorded postsynaptic responses will vary across repetitions of the same presynaptic stimulation. This variability is not just noise, but rather it carries important information about the synaptic dynamics. Accordingly, our approach seeks to select synaptic parameters so as to best describe the observed variability of the responses, instead of their averages. This is done by constructing a generative model of repetitive synaptic transmission, i.e. a parametric probability distribution function for the trains of postsynaptic responses given the times of presynaptic activation, and then determining the corresponding parameters by the maximum-likelihood principle.

We build a stochastic generative model of synaptic transmission by following a standard procedure which consists in augmenting the quantal model with dynamic processes that modulate the total release probability. According to the quantal model, a synaptic connection is a collection of N independent release sites. Upon the spike, a site can either release neurotransmitter or fail to do so. The probability of such an event (i.e., the total release probability) is decomposed into the product of the probability that the site is release competent and the probability that the release actually occurs, given that the site is release competent (Quastel, 1997). Hereafter, we refer to this latter simply as the release probability. A large class of stochastic models of synaptic transmission can be

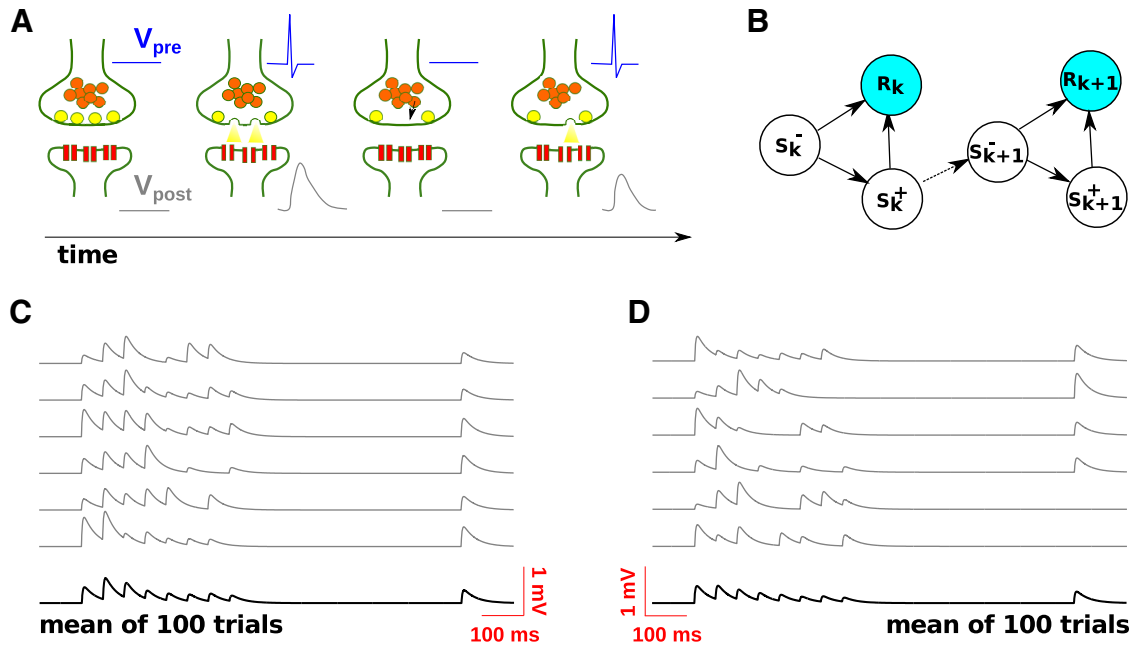


Figure 1. The generative model and sample synthetic traces. **A**, Schematics of the synaptic model. Upon spike (blue), only the docked vesicles (yellow) can be released. The postsynaptic response (gray) is proportional on average to the number of vesicles released. In between spikes, vesicles dock to noncompetent release sites (black arrow) with constant probability per unit time. **B**, Graphical model of the statistical dependencies among the number of docked vesicles before spike (S^-) and after spike (S^+) and observable responses (R). The number of docked vesicles is not observable. **C**, Sample synthetic traces and average trace for a facilitating connection: $N = 10$, $q = 0.15$ mV, $\sigma_q = 0.03$ mV, $U = 0.3$, $\tau_D = 195$ ms, $\tau_F = 570$ ms. **D**, Same as **C** for a depressing connection: $N = 10$, $q = 0.15$ mV, $\sigma_q = 0.03$ mV, $U = 0.25$, $\tau_D = 670$ ms, $\tau_F = 15$ ms.

formulated in this framework (indeed, all the models we are aware of), by appropriately selecting the dynamics of the release site and of the release probability (Fig. 1A). In the following, we chose the specific instantiations of those dynamics that correspond to the stochastic TM model (Fuhrmann et al., 2002).

In the stochastic TM model, the dynamics of the release site are described by the following simple docking process: a noncompetent site becomes competent with a constant probability per unit time $1/\tau_D$. A competent site becomes noncompetent by releasing upon spike. The probability of release $u(t)$ increases with each spike and decays back to its baseline level U , with a time constant τ_F in between spikes. The dynamics of $u(t)$ are a minimal phenomenological description of the effects of calcium influx into the synaptic terminal on the probability of release (Bertram et al., 1996; Markram et al., 1998; Dittman et al., 2000; Neher and Sakaba, 2008). The postsynaptic response to a single vesicle (quantal response) is variable with mean q (quantal size) and variance σ_q^2 (quantal noise). The postsynaptic response to multiple vesicles is simply the sum of the single quantal responses.

This model has two dynamic variables—the number of release-competent sites $S(t)$ and the release probability $u(t)$ —and six parameters—the number of release sites N , the quantal size q , the quantal noise σ_q^2 , the initial release probability U , and the time constants for docking and facilitation, τ_D and τ_F , respectively. It contains three sources of variability since both release and docking are stochastic processes, and the same number of released

vesicles can result in different postsynaptic responses due to quantal noise. These sources of variability are all well documented experimentally. In Figure 1, we show sample synthetic traces generated by the model, together with the trial-averaged trace, for a facilitating connection (Fig. 1C) and a depressing connection (Fig. 1D).

Having an explicit description for the sources of stochasticity, one can compute the probability that a given train of postsynaptic responses, $R_{1 \rightarrow M} \equiv \{R(t_1), R(t_2), \dots, R(t_M)\}$, is observed in correspondence with presynaptic spikes occurring at times $t_{1 \rightarrow M} \equiv \{t_1, t_2, \dots, t_M\}$. This probability as a function of the model parameters, which we collectively denote θ , is by definition the likelihood function, i.e.,

$$L(\theta | R_{1 \rightarrow M}) \equiv P(R_{1 \rightarrow M} | \theta). \tag{45}$$

The maximum-likelihood principle prescribes taking as an estimate of the parameters the values $\hat{\theta}_{ML}$, which maximize $L(\theta | R_{1 \rightarrow M})$. Although the likelihood function can be efficiently evaluated numerically (as we show in Materials and Methods), its direct maximization turns out to be impractical because the different responses are correlated, (i.e., $P(R_{1 \rightarrow M} | \theta) \neq \prod_{k=1}^M P(R_k | \theta)$). The origin of these correlations is easy to understand. Let us denote $S_{1 \rightarrow M}^- \equiv \{S(t_1^-), S(t_2^-), \dots, S(t_M^-)\}$ and $S_{1 \rightarrow M}^+ \equiv \{S(t_1^+), S(t_2^+), \dots, S(t_M^+)\}$ the number of release-competent sites immediately before and after the corresponding spikes, respectively. The probability of observing a given response, R_k , depends on the number of vesicles released

upon the k th spike; that is, it depends on both S_k^+ and S_k^- (in fact, it depends on their difference). Similarly, the probability of the next response being R_{k+1} depends on both S_{k+1}^+ and S_{k+1}^- . However, the probability of having a given number of release-competent sites S_{k+1}^- immediately before the $(k + 1)$ -th spike depends on the number of release-competent sites S_k^+ immediately after the k th spike (Fig. 1B). These probabilistic dependencies, which are graphically illustrated in Figure 1B, are the source of the correlations between the different synaptic responses. Note that, if properly dealt with, these correlations are a source of information about the underlying synaptic dynamics rather than a hindrance.

It should be clear from the above that the responses $R_{1 \rightarrow k}$ and $R_{k+1 \rightarrow M}$ are independent, conditionally on the knowledge of S_k^+ (Fig. 1B). Thus, the joint probability of the observed responses and the underlying sequence of the numbers of release-competent sites responsible for their generation can be conveniently factorized in the following way:

$$P(R_{1 \rightarrow M}, S_{1 \rightarrow M}^-, S_{1 \rightarrow M}^+ | \theta) = P(S_1^- | \theta) \prod_{k=1}^M P(S_k^+ | S_k^-, \theta) P(R_k | S_k^+, S_k^-, \theta) \prod_{k=1}^{M-1} P(S_{k+1}^- | S_k^+, \theta). \tag{46}$$

The conditional probabilities appearing in the above equation have the following straightforward interpretation: $P(S_1^- | \theta)$ is the initial distribution of the number of release-competent sites; $P(S_k^+ | S_k^-, \theta)$ is the probability that the number of release-competent sites changes from S_k^- to S_k^+ upon spike; $P(R_k | S_k^+, S_k^-, \theta)$ is the probability of observing a postsynaptic response R_k when the number of release-competent sites changes from S_k^- to S_k^+ ; and $P(S_{k+1}^- | S_k^+, \theta)$ is the probability that the number of release-competent sites changes from S_k^+ to S_{k+1}^- during the time interval $t_{k+1} - t_k$ in the absence of spikes. These conditional probabilities are easily computed from the model.

Unlike synaptic responses, however, the number of release-competent sites is not directly observable (i.e., is a hidden variable). A very powerful algorithm, the EM algorithm (Dempster et al., 1977; Dayan and Abbott, 2001), exists that allows a maximum-likelihood estimation in the presence of hidden variables, as it is our case. In Materials and Methods, we show how all the quantities needed to carry out EM can be efficiently computed, and we obtain explicit re-estimation formulas for the parameters.

Application to experimental data

Response variability

We began by analyzing response variability, which was not done in the study by Wang et al. (2006), and found that the synaptic responses exhibited strong variability. For the purpose of illustration, we show in Figure 2 sample voltage traces for one facilitating connection (Fig. 2A) and one depressing connection (Fig. 2B), together with the corresponding trial-averaged traces. In both cases, the large variability across the different repetitions is immedi-

ately evident. The variability is, indeed, so strong that the facilitating/depressing nature of the transmission is largely concealed in the single traces, while it becomes readily apparent in the trial-averaged traces.

To quantify variability, for each connection we extracted from the corresponding single-trial voltage traces the peak postsynaptic response corresponding to each presynaptic spike, as explained in Materials and Methods. For each connection, we then computed the coefficient of variation (CV) of each response, which was defined as the ratio between the SD of the response across the different trials and its average. The results of this analysis are shown in Figure 2C, where we report the histograms of the CV across the population of synaptic connections, separately for each response. The average of the initial response ranged between 0.11 and 3.43 mV (0.67 ± 0.58 mV; $n = 69$), while the associated CV ranged between 0.21 and 1.58 (0.59 ± 0.27 ; $n = 69$). These values are fully consistent with those from previous studies (Markram et al., 1997; Brémaud et al., 2007; Loebel et al., 2009). We took this as an indication of the reliability of the procedure we used for isolating synaptic responses within the single-trial voltage traces.

Synaptic unreliability remained high all along the stimulation, and it even increased for late responses, as can be seen in Figure 2C. This is a consequence of the increasing probability of failure due to vesicle depletion (Loebel et al., 2009). The population-averaged CV was smallest for the second and the recovery response. This is a consequence of the increasing probability of release occurring at facilitating synapses, while release-competent sites are still abundant (i.e., before depression builds up). The second and the recovery responses were, in fact, the most facilitated responses on average. Note that such high levels of variability are the rule, rather than the exception, for central chemical synapses (Markram et al., 1997; Brémaud et al., 2007; Loebel et al., 2009).

Two points are worth stressing. As we just discussed, changes in the level of variability of the responses carry information about the underlying synaptic dynamics. This information, however, is destroyed by the trial-averaging procedure needed to perform least-squares fitting. Second, the precision of the least-squares estimates for the parameters is fundamentally limited by the accuracy of the (experimental) estimates of the average responses. A straightforward application of the central limit theorem shows that, for a true CV of 0.3, an estimate of the average response within 10% relative precision would require about 80 repetitions, while an estimate within 5% relative precision would require >300 repetitions. Given the CVs estimated from the data, these figures should be considered as the minimal number of repetitions needed to estimate the average response with reasonable accuracy. Unfortunately, least-squares estimates can be grossly imprecise even when the empirical averages are determined with high accuracy, as we show in the Maximum-likelihood versus least-squares estimation section.

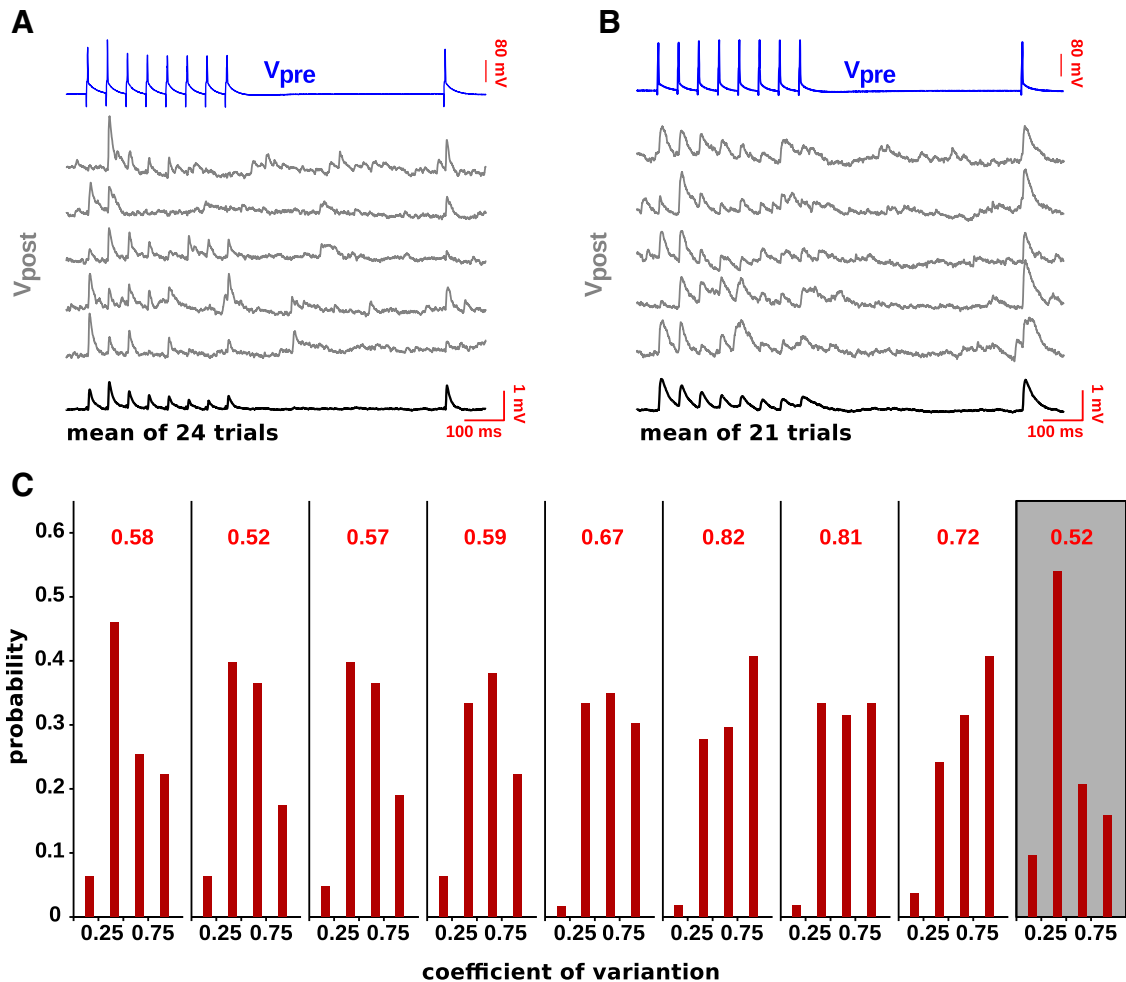


Figure 2. Stimulation protocol and variability of synaptic responses (experimental data). **A**, Five sample single-trial trains of experimentally measured postsynaptic responses (gray traces) illustrate the large trial-to-trial variability. The trial-averaged trace (black trace) reveals facilitating transmission (compare first and second response). Stimulation consists of a regular train of eight spikes followed by a recovery spike (top blue trace). **B**, Same as in **A** for a depressing connection (compare first and second responses). **C**, Histograms of the CVs of the synaptic responses in the train and upon the recovery spike (gray-shaded panel) across the dataset. The red numbers denote the respective average CVs. The size of the bins is 0.25, apart from the last one, which includes all CVs >0.75.

Maximum-likelihood estimation of the synaptic parameters

The synaptic parameters $\theta = \{N, q, \sigma_q, U, \tau_D, \tau_F\}$ for a given connection were estimated as follows. For a fixed value of N (note that N takes on only integer values, while the other parameters are continuous), the maximum-likelihood estimate of the remaining parameters is obtained by using the EM algorithm (see Materials and Methods). We varied N between 1 and 100, and the value of N and of the corresponding continuous parameters for which the likelihood was maximal was then selected as the final estimate.

The procedure is illustrated in Figure 3A for a sample connection. In Figure 3A, left, we plot the log-likelihood as a function of N , while, in Figure 3A, right panels, we plot the values of the parameters that maximize the log-likelihood for the corresponding N . The log-likelihood exhibits a clear maximum at $N = 17$. The values of the

remaining parameters can be read from the corresponding curves on the right. They are as follows: $q = 0.18$ mV, $\sigma_q = 0.06$ mV, $U = 0.27$, $\tau_D = 202$ ms, and $\tau_F = 449$ ms. Using these parameters, we generated 500 synthetic experiments in which the model was probed with the same stimulation protocol, and for the same number of trials (28), as in the real experiment. We then computed the average responses and the associated CVs for each experiment, and from these the corresponding grand averages together with 95% confidence intervals. The results are shown in Figure 3B. In the top panel of Figure 3B, we report the experimental (black curve) and the model average responses (red curve; error bars represent 95% confidence intervals). In the bottom panel of Figure 3B, we report the experimental (black curve) and the synthetic CVs (red curve; error bars represent 95% confidence interval). The model parameters were not selected to reproduce the average responses or the CVs, but rather to

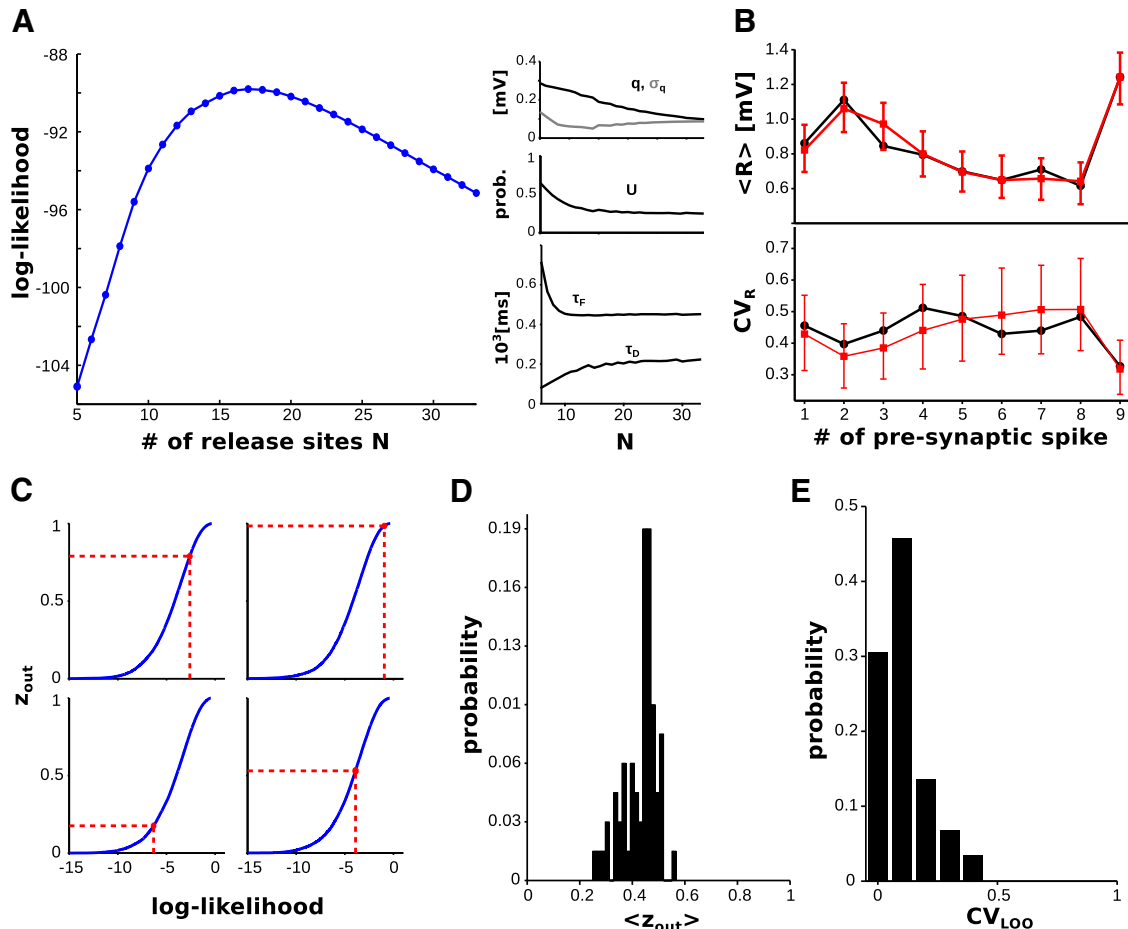


Figure 3. Maximum-likelihood estimation of the synaptic parameters from experimental data. **A**, Log-likelihood (left) and associated model parameters (right) as a function of N for a sample connection. The maximum is attained at $N = 17$ with $q = 0.18$ mV, $\sigma_q = 0.06$ mV, $U = 0.27$, $\tau_D = 202$ ms, $\tau_F = 449$ ms. **B**, Top, Average experimental responses (black line) vs average model responses (red line) for the same connection as in **A**. Bottom, Same as in the top panel for the coefficients of variation. Error bars indicate the 95% confidence interval of the model prediction. **C**, Cumulative distribution function of the log-likelihood for four instances of the leave-one-out procedure (blue curves). The red dots indicate the log-likelihood of the left-out trials. Same connection as in **A**. **D**, Distribution over the dataset of the average z_{out} . **E**, Distribution over the dataset of the average coefficients of variation of the estimates obtained in the leave-one-out procedure.

maximize the probability of the occurrence of the actual trains of responses observed in the experiment. Nevertheless, as can be seen, the experimental data are well reproduced by the model.

We estimated the synaptic parameters for all the connections in our dataset. In 6 of 69 cases, the estimation procedure returned values for one or more parameters that were judged to be problematic. In four cases, the estimation procedure returned values for one or both the time constants (i.e., τ_D and τ_F) that were several orders of magnitude larger than the longest timescale at which the synaptic connections were probed. In the two remaining cases, there was no maximum in the range of N values probed. These connections were excluded from further analysis.

We checked next whether the model was overfitting the data by using a standard leave-one-out cross-validation procedure (Stone, 1974). We estimated the parameters as described above while leaving out one trial every time.

With the parameters obtained, we computed the probability that the model would generate a set of responses with a log-likelihood equal or smaller than the log-likelihood of the set of responses that was left out. Hereafter, we denote this probability by z_{out} . This procedure is illustrated in Figure 3C for the same connection as in Figure 3A. In each subpanel, we show (1) the cumulative distribution of the log-likelihood for a set of responses generated from the model, where the parameters are estimated by leaving out one trial (blue curve); (2) the value of the log-likelihood of the set of responses left out during the estimation procedure (on the x-axis); and (3) the corresponding value of z_{out} (on the y-axis). For sets of responses generated by the model, one expects z_{out} to be uniformly distributed between 0 and 1. The distribution of the average (over all trials) z_{out} value across the dataset is shown in Figure 3D. As can be seen, most of the values are between 0.4 and 0.5. For only 5 of 63 connections, the distribution of z_{out} values obtained from the leave-one-out

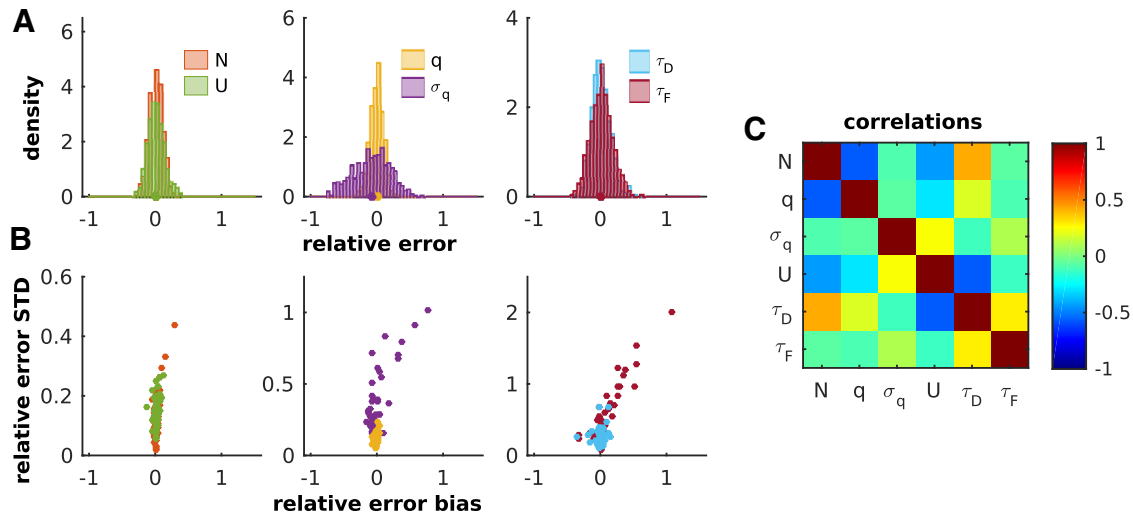


Figure 4. Estimating uncertainty and correlations between parameter estimates with parametric bootstrap. **A**, Distributions of the relative errors obtained by re-estimating the parameters of the sample connection in Fig. 3A from synthetically generated responses. The dots on the x-axes indicate the corresponding averages. **B**, Standard error vs bias of the relative error for all the connections. **C**, Pearson correlation coefficients between all pairs of parameter estimates averaged over all the connections.

procedure was statistically different from the uniform distribution (Kolmogorov–Smirnov test, $p = 0.01$). For three of these connections, <25 trials were available. No connection with ≥ 30 trials ($n = 10$) showed a distribution of z_{out} values that was statistically different from the uniform distribution.

The leave-one-out procedure also allowed us to evaluate the stability of the estimation procedure, which we quantified by computing the average coefficient of variation of the estimates. The distribution obtained is reported in Figure 3E.

We concluded that, for the large majority of the connections (58 of 63), there were no indications of overfitting, while for the remaining connections it was likely that we simply lacked statistical power to assess overfitting. Also our estimation procedure exhibited an accuracy and stability that tended to be better than the least-squares fitting procedure (see the Maximum-likelihood versus least-squares estimation section), while allowing us to estimate two additional parameters.

Quantifying the uncertainty of the parameter estimates

We used a standard re-sampling procedure (i.e., parametric bootstrap; Efron and Tibshirani, 1994) to quantify the uncertainty of the parameter estimates. For each set of parameters obtained from a synaptic connection in our dataset, we generated 500 synthetic experiments with the same settings as in the original experiment (i.e., same stimulation protocol, number of repetitions, and baseline noise). We then re-estimated the parameters for each synthetic experiment and computed their relative errors with respect to the original set of parameters. We show in Figure 4A the resulting distributions of relative errors for the same connection shown in Figure 3A. All the distributions peaked at ~ 0 , with averages very close to 0 and relative SDs < 0.3 . This shows that, at least for this con-

nection, our method returns unbiased and accurate estimates of the parameters.

We see in Figure 4B, where we plot the SD versus the bias of the relative error for each parameter and each connection, that this is, in fact, the case for the majority of connections in our dataset. For some connections, however, the estimates of σ_q and/or τ_F are quite inaccurate. Large relative errors in the estimates of σ_q are due to the connections for which the baseline noise is comparable or larger than the quantal noise. Large relative errors in the estimates of τ_F are due to connections for which the (experimentally) chosen interstimulus interval was ill suited to resolve the facilitation time course. For instance, in depression-dominated connections (i.e., $\tau_F < \tau_D$) the facilitation time constant cannot be reliably estimated if the interstimulus interval is much longer than τ_F . The high uncertainty in the estimates of σ_q and τ_F (for some connections) is responsible for the positive biases observed in Figure 4B. This is because relative errors cannot be smaller than -1 , given that both parameters have to be positive, while they can be quite large and positive (i.e., $\gg 1$). It is important to stress that the large variability and the accompanying bias in the estimates of σ_q and τ_F are not due to the method itself but to the small amount of information conveyed by the data (experimental or synthetic). This can be remedied, especially for τ_F , by choosing more informative stimulation protocols (see the Estimating synaptic parameters from a single spike train section).

Using the data generated in the synthetic experiments, we also computed, for each connection, the (Pearson) correlation coefficients between all pairs of parameter estimates. The correlation coefficients were then Fisher transformed (Efron and Tibshirani, 1994), averaged across connections, and Fisher transformed back. The result is shown in Figure 4C. Fluctuations in the parameter estimates are all significantly correlated. This is to be ex-

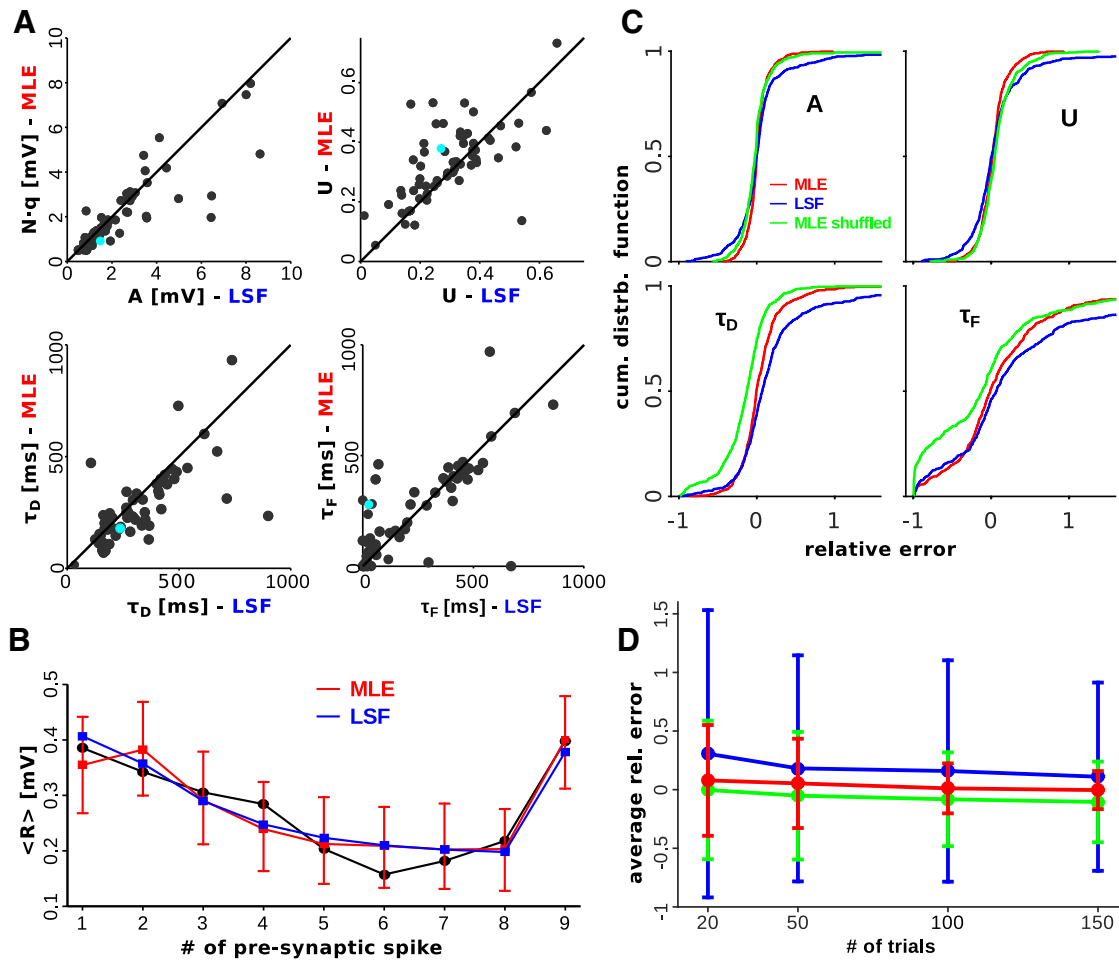


Figure 5. Maximum-likelihood estimation (MLE) vs least-squares fitting (LSF). **A**, Estimates obtained with MLE vs those obtained with LSF for all the connections in the dataset. **B**, Average experimental responses (black line) for the connection corresponding to the cyan dot in **A** together with the result of LSF (blue line) and the MLE prediction with 95% confidence interval (red line + bars). MLE estimate: $A = 0.91$ mV, $U = 0.38$, $\tau_D = 179$ ms, $\tau_F = 279$ ms; LSF estimate: $A = 1.48$ mV, $U = 0.27$, $\tau_D = 236$ ms, $\tau_F = 28$ ms. **C**, Cumulative distribution functions of the relative errors of MLE (red), MLE with shuffled responses (green) and LSF (blue) estimates from synthetic experiments. **D**, Average and SD (bars) of relative errors for MLE (red), MLE with shuffled responses (green), and LSF (blue) estimates as a function of the number of trials (synthetic experiments).

pected. An estimate that deviates from its true value necessarily induces compensatory adjustments in one or more of the other parameters. For instance, the overestimation of N must be compensated by the underestimation of q (or U or both) in order to reproduce the empirically observed range of the responses. In fact, correlations between N and q , N and U , and U and q were all negative and quite substantial ($R = -0.56$, $R = -0.44$, and $R = -0.32$, respectively). Similarly, the underestimation of N must be compensated by the underestimation of τ_D (i.e., faster recovery from depression) in order to maintain the empirically observed range of responses throughout the train. Accordingly, N and τ_D are positively correlated ($R = 0.40$).

Maximum-likelihood estimation versus least-squares fitting

Standard least-squares fitting allows one to estimate the initial release probability U , the two time constants τ_D and

τ_F , and the product $N \cdot q$, which we denote by A (for details, see Materials and Methods). We thus compared the estimates obtained with our method with the ones obtained by least-squares fitting. Note that the two methods would return asymptotically the same estimates for these parameters. In **Figure 5A**, we plot the estimates obtained with our method versus the estimates obtained with the least-squares fitting procedure. As can be seen, the two methods tend to return correlated estimates. Nevertheless, for more than half of the connections, the relative difference between the estimates with the two methods was $>30\%$ for at least one of the parameters. In several cases, the two estimates were dramatically different. One such case is illustrated in **Figure 5B**. According to maximum-likelihood estimation, this connection was facilitating ($\tau_F = 279$ ms, $\tau_D = 179$ ms, $\tau_F/\tau_D = 1.56$), while according to least-squares fitting the same connection was depressing ($\tau_F = 28$ ms, $\tau_D = 236$ ms, $\tau_F/\tau_D = 0.12$). Note that the estimates of τ_F differed by one order of

magnitude. The large differences between the estimates with the two methods thus suggest that, for most of the connections, the number of repetitions is too small to reach the asymptotic regime.

Our method makes use of more information (i.e., correlations and variability) to estimate the parameters. One would accordingly expect more accurate estimates compared with least-squares fitting, and the often quite large discrepancies would thus result from the different accuracies of the two methods. To test this hypothesis, we generated 500 synthetic connections by randomly and independently selecting their parameters from the corresponding experimental distributions (as estimated with our method; see Population analysis section). The synthetic connections were probed with a regular train of eight spikes at 20 Hz, followed by a recovery spike 550 ms after the end of the train. We collected the responses over 20 trials for each connection, and re-estimated its synaptic parameters. To investigate the impact of the correlations on the accuracy of the estimates, we also applied our method to shuffled response trains. These were obtained by randomly reassigning responses to trains while preserving their position within the train (e.g., the second response in the first train becomes the second response in the fifth train). The distributions of the relative errors, with respect to the true parameters, obtained with our method, with shuffled trains and least-squares fitting, are shown in [Figure 5C](#). As expected, the estimates obtained by least-squares fitting as well as those obtained from the shuffled trains were more inaccurate than those obtained with our method. Note that the shuffling procedure introduces a systematic error in both the estimate of τ_D and the estimate of τ_F . This is because consecutive responses are negatively correlated when release can occur from a finite number of sites. This correlation decays exponentially over a timescale of the order of the docking time. The shuffling procedure tends to reduce the correlation between consecutive responses, which results in the systematic underestimation of τ_D . This, in turn, leads to the underestimation of τ_F , as the two estimates tend to be positively correlated ([Fig. 4C](#)).

We next investigated how the accuracy of the estimates increased with the number of trials. We repeated the analysis described above when collecting responses over 50, 100, and 150 trials. We then computed the mean and the SD of the corresponding distributions of the relative errors ([Fig. 5C](#)). The results are plotted in [Figure 5D](#) (our method, red curve; our method with shuffled trains, green curve; least-squares fitting, blue curve). For the estimates obtained with our method, the mean of the relative errors, which is already very small for 20 trials, quickly converges to 0. Likewise, the accuracy of the estimates, as measured by the SD of the relative errors ([Fig. 5D](#), bars), steadily increases with the number of trials. The estimates obtained from shuffled trains and those obtained by least-squares fitting exhibited the same overall trends. Their accuracy, however, improved in a much slower way with an increasing number of trials. Note that, even for 150 trials, the accuracy was significantly smaller than the accuracy of the estimates obtained with our method.

This is particularly evident, and puzzling, for the estimates obtained by least-squares fitting. The slow increase in the accuracy of the estimates with the number of trials could be a consequence of the fact that the least-squares procedure is more prone than the maximum-likelihood procedure to get stuck in local minima (as suggested by [Costa et al., 2013](#)). Alternatively, it could result from the fact that the least-squares procedure is ill posed. To understand which of these two possibilities was the more likely explanation of the observed phenomenology, we computed, for each connection in our synthetic sample (see above), the condition number (for details, see Materials and Methods) and estimated the range of the relative errors obtained by the least-squares procedure when adding a small amount of noise to the true average responses. Concretely, for each connection we added Gaussian noise with a relative SD of 0.01 independently to each true average response in the train, and then re-estimated the parameters. We repeated this procedure 15 times, and took the largest difference between relative errors (of the same parameter) as an estimate of the corresponding range. In [Figure 6A](#), we plot the range of relative errors versus the condition number. As can be seen, for a fraction of the connections the condition number, and correspondingly the range of relative errors, is quite large. For these connections, we verified that there was no global minimum in the neighborhood of the true parameters, contrary to what one would expect for a well posed least-squares procedure. Thus, large relative errors were not a result of the minimization routine getting stuck in local minima. For the purpose of illustration, we report in [Figure 6, B and C](#), the samples least-squares fit to two connections for which the condition number was >100 . As can be seen, the estimated parameters fluctuate substantially across the different noise realizations.

Although we cannot rule out definitively the possibility that the least-squares procedure returning local minima contributes to the poor increase in the accuracy of the estimates with the number of trials, we believe that that is largely a result of the fact that, for some set of parameters, the least-squares fitting is very sensitive to the specific realizations of the synaptic responses. This could happen, as we have just shown, even for experimentally unrealistic large number of trials; to achieve a 1% relative error on the average, one would need a number of trials on the order of 10^4 .

Population analysis

The average values and ranges of the six synaptic parameters (N , q , σ_q , U , τ_D , τ_F) obtained with our method are reported in [Table 1](#), along with the estimates obtained with other methods in similar preparations and conditions ([Larkman et al., 1991, 1997](#); [Markram et al., 1998](#); [Silver et al., 2003](#); [Koester and Johnston, 2005](#); [Wang et al., 2006](#); [Brémaud et al., 2007](#); [Loebel et al., 2009](#); [Hardingham et al., 2010](#)). The corresponding distributions are shown in [Figure 7A](#). It is important to stress that our method returned quantal parameters in excellent agreement with published estimates, although the number of

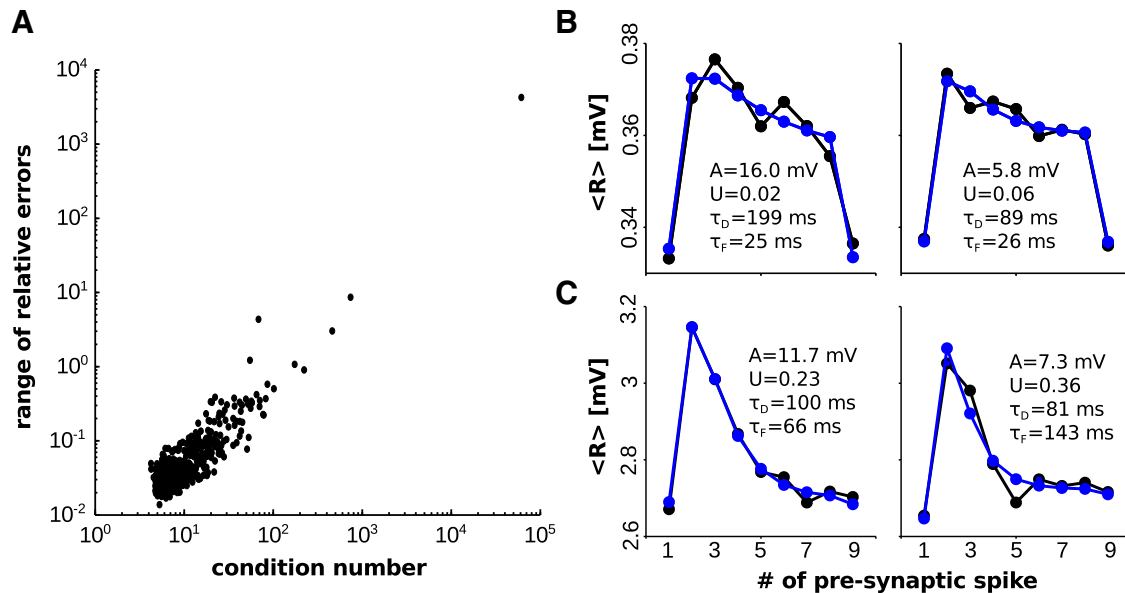


Figure 6. The condition number and the accuracy of the least-squares estimates. **A**, The range of relative errors vs the condition number of the estimates obtained by least-squares fitting (synthetic experiments). **B**, Sample synthetic average responses (black) together with average model responses (blue) resulting from the least-squares fitting. The parameters reported in the panels are the least-squares estimates. True parameters were as follows: $A = 4.8$ mV, $U = 0.07$, $\tau_D = 95$ ms, $\tau_F = 28$ ms. **C**, Same as in **B** but with true parameters: $A = 8.1$ mV, $U = 0.33$, $\tau_D = 81$ ms, $\tau_F = 100$ ms.

observed responses available for each connection was several times smaller than the number of observed responses typically required for current state-of-the-art methods (Silver, 2003).

The average first responses (i.e., the synaptic efficacies as routinely estimated) varied over a 32-fold range (from 0.11 to 3.43 mV) across connections. Their distribution was very well fitted by a log-normal distribution (one-sample Kolmogorov–Smirnov test, $p \approx 0.74$; log location, -0.65 ; log scale, 0.73), consistent with what was reported for other cortical regions (Song et al., 2005; Loewenstein et al., 2011; Buzsáki and Mizuseki, 2014). To investigate the origin of such a large range, we searched for correlations between the synaptic efficacy and the quantal parameters N , q , and U . We found a very strong correlation with the number of release sites N ($R = 0.76$, $p < 10^{-13}$; Fig. 7B), and a weaker but significant correlation with the quantal amplitude q ($R = 0.25$, $p < 0.05$). We found no correlation with the initial release probability U . Also, we did not find significant correlations among these three parameters. We concluded that synaptic efficacy is primarily determined by the number of release sites, which is consistent with previous reports (Loebel et al., 2009; Chabrol et al., 2015). Correlation between synaptic efficacy and quantal amplitude has also been reported previously (Hardingham et al., 2010).

Next, we searched for correlations between synaptic parameters. The initial release probability U was negatively correlated with both the time constant of the docking process τ_D ($R = -0.48$, $p < 10^{-4}$; Fig. 7C) and the time constant for facilitation τ_F ($R = -0.34$, $p < 10^{-3}$; Fig. 7D). It has been shown that the total number of vesicles released during a train of action potentials increases with

the initial release probability (Dobrunz and Stevens, 1997). This suggests that the larger the initial release probability, the faster the docking process, which is consistent with the negative correlation between U and τ_D that we found. Similarly, large paired-pulse ratios (i.e., strong facilitation) are typically observed in synaptic connections with a low initial release probability (Magleby, 1979; Dobrunz and Stevens, 1997; Oleskevich et al., 2000). Again, this is consistent with the negative correlation between U and τ_F that we found. The same trends were also found with the estimates obtained with least-squares fitting. However, the correlation coefficients were weaker, and less statistically significant ($R = -0.34$, $p = 0.007$ for the correlation between U and τ_D ; $R = -0.24$, $p = 0.06$ for the correlation between U and τ_F). We did not find other pairs of synaptic parameters exhibiting statistically significant correlation.

Estimating synaptic parameters from a single spike train

In this section, we demonstrate one of the main advantages of our method over existing state-of-the-art methods. The least-squares fitting procedure requires the stimulation to be broken into separate trials and these trials to be repeated for putatively identical initial conditions, which is typically obtained by introducing quite long intervals (4–10 s) between consecutive trials. The method we have developed is free from both of these constraints. To illustrate the advantages this fact entails, we generated a synthetic connection by setting the parameters to the corresponding experimental population-averaged values, apart from N , which we set to the mode of the experimental distribution (i.e., $N = 8$), and probed it with three different stimulation protocols. The regular protocol is the

Table 1. Parameter estimates obtained with our method (in bold type) compared with published estimates obtained in similar experimental preparations

Parameter	Average ± SD	Range	Preparation	Temperature (°C)	Method	Reference
	15 ± 12	2–77	Ferret mPFC, L5 to L5	32–34	MLE	
	8.1 ± 5.3	3–18	Rat HC, CA1	34	H, QMLE	Larkman et al. (1997)
	5.1 ± 2.7	2.3–11.1	Rat SC, L4 spiny stellate to L2/3 pyramidal	35–37	MV	Silver et al. (2003)
<i>N</i>	7.48 ± 7.11	2–22	Rat VC and SC, various layers	35–36	MV, FA	Brémaud et al. (2007)
	12.21 ± 7.9	2–32	Cat VC, various layers	35–36	MV, FA	Brémaud et al. (2007)
	53.3 ± 42	7–170	Rat SC, L5 to L5	35	LS	Loebel et al. (2009)
	3.4 ± 2.2	1–7	Rat VC, L5 to L5	23–26/36	MV, H	Hardingham et al. (2010)
	0.15 ± 0.06	0.06–0.32	Ferret mPFC, L5 to L5	32–34	MLE	
<i>q</i> [mV]	0.131 ± 0.145	0.084–0.197	Rat HC, CA1	34	H, QMLE	Larkman et al. (1991)
	0.196 ± 0.062	0.066–0.275	Rat HC, CA1	34	H, QMLE	Larkman et al. (1997)
	0.15 ± 0.09		Rat SC, L4 spiny stellate to L2/3 pyramidal	35–37	MV	Silver et al. (2003)
	0.37 ± 0.17	0.15–1	Rat VC and SC, various layers	35–36	MV, FA	Brémaud et al. (2007)
	0.22 ± 0.19	0.07–1	Cat VC, various layers	35–36	MV, FA	Brémaud et al. (2007)
	0.13 ± 0.04	0.06–0.25	Rat SC, L5 to L5	35	LS	Loebel et al. (2009)
	0.211 ± 0.065	0.106–0.302	Rat VC, L5 to L5	23–26/36	MV, H	Hardingham et al. (2010)
	0.05 ± 0.05	0.00–0.31	Ferret mPFC, L5 to L5	32–34	MLE	
σ_q [mV]	0.052 ± 0.085	0.030–0.10	Rat HC, CA1	34	H, QMLE	Larkman et al. (1991)
	0.065 ± 0.15		Rat SC, L4 spiny stellate to L2/3 pyramidal	35–37	MV	Silver et al. (2003)
	0.33 ± 0.13	0.05–0.73	Ferret mPFC, L5 to L5	32–34	MLE	
	0.53 ± 0.17	0.14–0.81	Rat HC, CA1	34	H, QMLE	Larkman et al. (1997)
	0.59 ± 0.16 [†]		Rat SC, L5 to L5	32–34	LS	Markram et al. (1998)
	0.05 ± 0.04 [‡]		Rat SC, L5 to L5	32–34	LS	Markram et al. (1998)
	0.79 ± 0.12		Rat SC, L4 spiny stellate to L2/3 pyramidal	35–37	MV	Silver et al. (2003)
<i>U</i>	0.46 ± 0.26*	0.15–0.95	Rat SC, L2/3 to L2/3	35	OQA	Koester and Johnston (2005)
	0.27 ± 0.15	0.03–0.6	Ferret mPFC, L5 to L5	32–34	LS	Wang et al. (2006)
	0.63 ± 0.6		Rat VC and SC, various layers	35–36	MV, FA	Brémaud et al. (2007)
	0.69 ± 0.18		Cat VC, various layers	35–36	MV, FA	Brémaud et al. (2007)
	0.46 ± 0.1	0.25–0.65	Rat SC, L5 to L5	35	LS	Loebel et al. (2009)
	0.46 ± 0.21	0.15–0.75	Rat VC, L5 to L5	23–26/36	MV, H	Hardingham et al. (2010)
	0.53 ± 0.05		Rat VC, L5 to L5	32–34	Bayesian	Costa et al. (2013)
	335 ± 306	16–1800	Ferret mPFC, L5 to L5	32–34	MLE	
τ_D [ms]	813 ± 240 [†]		Rat SC, L5 to L5	32–34	LS	Markram et al. (1998)
	399 ± 295 [‡]		Rat SC, L5 to L5	32–34	LS	Markram et al. (1998)
	396 ± 163	0–1600	Ferret mPFC, L5 to L5	32–34	LS	Wang et al. (2006)
	525 ± 134	380–900	Rat SC, L5 to L5	35	LS	Loebel et al. (2009)
	321 ± 340	0–1900	Ferret mPFC, L5 to L5	32–34	MLE	
τ_F [ms]	1797 ± 1247 [‡]		Rat SC, L5 to L5	32–34	LS	Markram et al. (1998)
	292 ± 240	0–1600	Ferret mPFC, L5 to L5	32–34	LS	Wang et al. (2006)

The last three columns display the temperature at which recordings were taken, the estimation methods, and the reference to the corresponding study, respectively. All synaptic connections are pyramidal to pyramidal, unless stated otherwise. HC, Hippocampus; VC, visual cortex; SC, somato-sensory cortex; L, layer; MV, mean-variance analysis; QMLE, quantal maximum likelihood; LS, least-squares fit; H, histogram; FA, failure analysis; OQA, optical quantal analysis. *Release probability per connection; [†]dominantly depressing connections; [‡]dominantly facilitating connections.

same as the experimental protocol used to collect the data analyzed in this study (Fig. 2). The Poisson protocol is obtained from the regular protocol, substituting in each trial the regular train with a Poisson train with the same average interspike interval. In other words, in the Poisson protocol the spike train is different on each trial, thus removing the constraint of the repetition of the same stimulation across trials. In both the regular and Poisson protocol consecutive trials are separated by a 4 s interval.

The single-sweep protocol is obtained from the Poisson protocol by removing the intertrial interval, i.e., the recovery spike of the previous trial coincides with the first spike of the next trial (Fig. 8A). The single-sweep protocol removes the constraint of repeating the trials for putatively identical initial conditions.

We collected responses for the same recording time (~2 min) with the three protocols described above. This corresponds to the same number of responses (on aver-

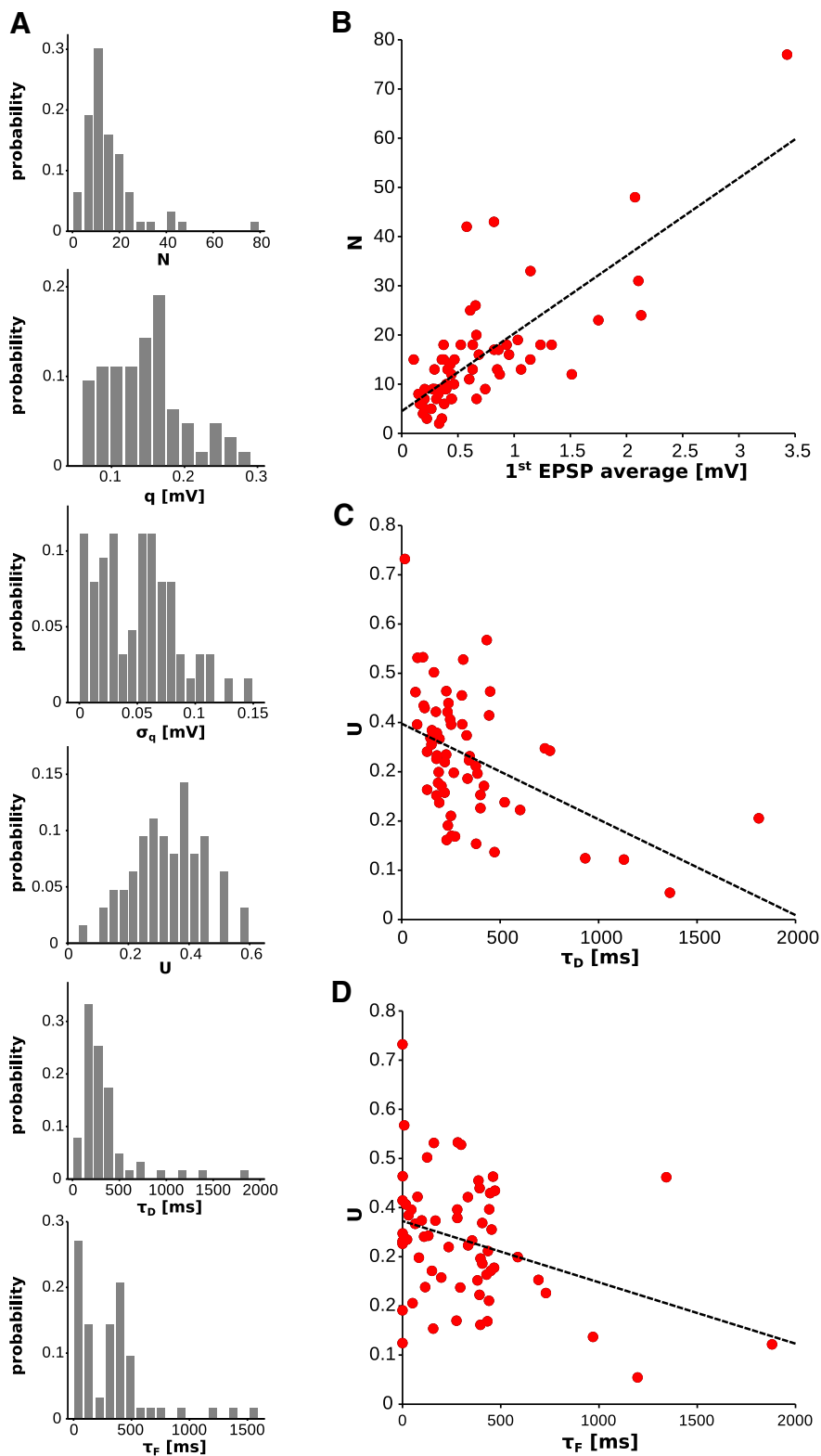


Figure 7. Distributions and correlations of the parameters (experimental data). **A**, Distributions over the dataset of the different synaptic parameters. **B**, Number of release sites N vs average first response in the train. Dashed line, Linear regression ($R = 0.76$, $p < 10^{-13}$). **C**, Initial release probability U vs time constant of the docking process τ_D . Dashed line, Linear regression ($R = -0.48$, $p < 10^{-4}$). **D**, Initial release probability U vs time constant of facilitation τ_F . Dashed line, Linear regression ($R = -0.34$, $p < 10^{-3}$).

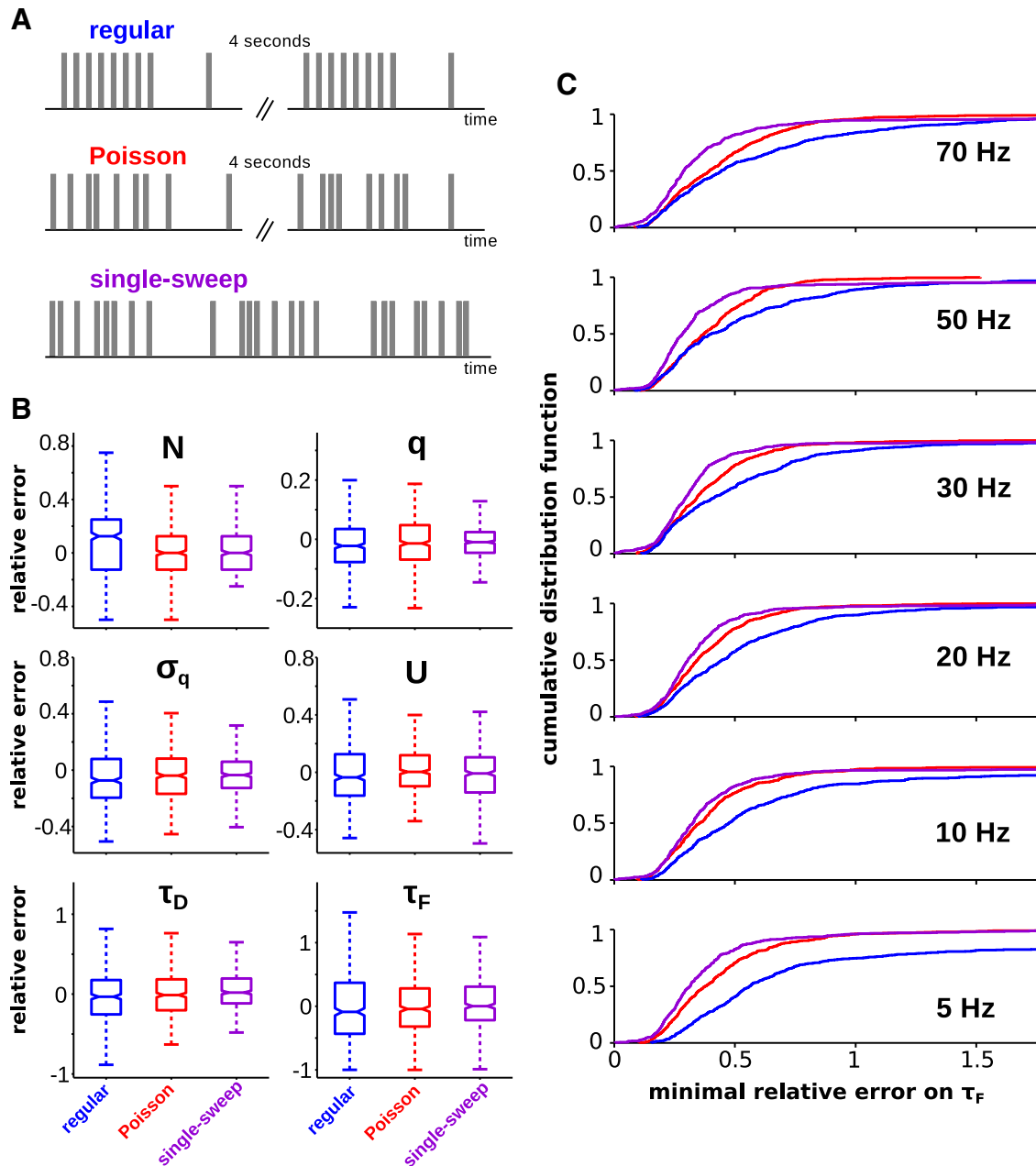


Figure 8. Comparison of different stimulation protocols (synthetic experiments). **A**, Schematics of the protocols. **B**, Box plots of the distributions of the relative errors of the estimates for all the parameters for sample synthetic connection. Stimulation frequency was 5 Hz. **C**, Cumulative distribution function of the minimal relative errors on the estimate of τ_F at varying stimulation frequencies for the regular (blue lines), the Poisson (red lines) and the single-sweep (magenta lines) protocols. Cumulative distribution functions at the same stimulation frequency are all statistically different (Kolmogorov–Smirnov test, $p < 0.01$). This is true for all stimulation frequencies.

age) for the regular and the Poisson protocols (180 responses with the spike trains at 5 Hz), and, clearly, to a larger number of responses for the single-sweep protocol (440 responses on average, again with spike trains at 5 Hz). From the data so collected, we estimated the parameters of the synthetic connection with our method. The resulting distributions of relative errors (over 500 synthetic experiments) are shown in Figure 8B. Several points are noteworthy. First, the regular protocol appears to be the

least effective one. Both the Poisson and the single-sweep protocol (Fig. 8B, red and purple box plots, respectively) led to estimates that are significantly more accurate than the estimates obtained with the regular protocol (blue box plots). Especially interesting is the fact that the Poisson protocol outperforms the regular one at parity of number of responses. This is because the trains of responses generated by the Poisson protocol are more variable, with the additional variability coming from the

stochasticity of the presynaptic spike times. The stochasticity of the stimulation forces the synapse to visit a larger fraction of its possible hidden states and, consequently, to produce more informative sequences of observable responses (see below). Second, the single-sweep protocol outperforms the Poisson protocol. This is a consequence of the larger number of responses collected. At parity for the number of responses, the Poisson protocol produced more accurate estimates than the single-sweep estimate (data not shown). The reason is that, after a long intertrial interval (i.e., much longer than for τ_D and τ_F), the hidden state is known with very high probability. This, clearly, facilitates the estimation of the synaptic parameters. The increase in the number of responses collected during the same recording period led to more accurate estimates nonetheless.

The advantage of using the single-sweep protocol we have just illustrated does not depend on the specific parameters that were chosen for the synthetic connection or on the specific stimulation frequency. To show this, we calculated the FIM for randomly generated synthetic connections (as described in the Maximum-likelihood estimation versus least-squares fitting section) as a function of the stimulation protocol (i.e., regular, Poisson, and single sweep) and of the stimulation frequency. The FIM allows one to lower bound the variance that an unbiased estimator of the (continuous) synaptic parameters can achieve for a given protocol of stimulation. In other words, the FIM provides a precise measure of how informative the responses collected with the given protocol are about the parameters. The procedure followed to compute the FIM is described in Materials and Methods. Using the lower bound from the FIM, we computed the minimal relative errors achievable for all (continuous) synaptic parameters.

As an illustration, we plot in [Figure 8C](#) the cumulative distributions of these minimal relative errors for the facilitation time constant, for varying stimulation protocols and frequencies (recording time is 118 s). We chose to show the relative errors of the facilitation time constant because it is the parameter that is hardest to estimate, according to the FIM analysis. As can be seen, the single-sweep protocol ([Fig. 8C](#), purple curves) outperforms the Poisson protocol ([Fig. 8C](#), red curves) which, in turn, outperforms the regular protocol ([Fig. 8C](#), blue curves) independent of the stimulation frequency. It is worth noting that the accuracy of the estimates is strongly dependent on the stimulation frequency for the regular protocol. On the other hand, this dependency is much weaker for both the Poisson and the single-sweep protocol. This can be intuitively understood by noticing that, with the regular protocol, the synaptic connection is probed only on two timescales (i.e., the interspike interval of the train and the recovery interval). If these two timescales are much longer than the synaptic dynamics timescales, these latter ones can only be poorly resolved. By contrast, the other protocols probe the dynamics with widely different interspike intervals (i.e., they are exponentially distributed).

Discussion

We have introduced a new framework to quantitatively characterize patterns of transmission at chemical synapses in a statistically principled way. By explicitly modeling the sources of stochasticity present in the process, we obtained a parametric description of the probability of observing a sequence of postsynaptic responses upon a given pattern of presynaptic activation. The estimate is then obtained by selecting the parameters that maximize the likelihood of the observed synaptic responses. The uncertainty and the bias of the estimate are obtained by parametric bootstrap. If prior knowledge about the synaptic parameters is available, the whole procedure can be straightforwardly transformed into maximum a posteriori estimation. As one of the main sources of stochasticity stems from the quantal release of the neurotransmitter, our method naturally estimates the quantal parameters along with the dynamic ones from the same set of synaptic responses. Our method significantly outperforms the standard least-squares fitting procedure in terms of the accuracy of the estimates, irrespective of the available number of trials. This is because, unlike least-squares fitting, our method is able to extract information about the parameters contained in the correlation between synaptic responses and in their variability. When applied to experimental data, our method returns estimates of the parameters that are in excellent agreement with the published literature, despite the small number of synaptic responses (~ 100) used and the larger number of parameters being estimated ([Silver, 2003](#)).

Generality and robustness of the method

In this study, we have chosen the stochastic TM model as the underlying generative model ([Fuhmann et al., 2002](#)). Several considerations motivated this choice: (1) the TM model has a small number of free parameters and yet is able to describe very diverse patterns of transmission; (2) The dataset we have analyzed was previously analyzed with the TM model, which allows for direct comparison; and (3) the stochastic TM model has been widely used in theoretical investigations of functional/computational implications of synaptic variability; thus, it appeared relevant to assess to which extent the model captures variability in real synapses. Our method, however, is general and can be extended to accommodate more biophysical detail, such as calcium-dependent docking rates ([Dittman et al., 2000](#)) or the presence of multiple vesicle pools ([Rizzoli and Betz, 2005](#)). Also, we emphasize that our method is not restricted to electrophysiological recordings and could be adapted to optical recordings ([Yuste et al., 1999](#); [Oertner et al., 2002](#); [Trigo et al., 2012](#)).

The frequencies used to probe synaptic transmission in our dataset were relatively low, and, accordingly, we have neglected postsynaptic receptor desensitization. However, desensitization is known to significantly contribute to short-term depression at higher stimulation frequencies ([Trigo et al., 2012](#)). This is especially relevant when experimentally probing synaptic connections with the Poisson protocol where interstimulus intervals can be very short. Our framework can be extended to include desensitiza-

tion by making the quantal size dependent on spiking history. Explicit re-estimation formulas can be obtained and evaluated numerically in an efficient way also in this case (A. Barri and D.A. DiGregorio, unpublished observations).

Similarly, we have neglected heterogeneities across the release sites (Silver, 2003). Numerical experiments (data not shown) suggest that weak-to-moderate heterogeneities (coefficient of variation between 0.25 and 0.5) in the release site parameters are effectively absorbed in the estimate of the quantal noise. The estimates of the other parameters are very close to the averages of the corresponding distributions. The estimate of the number of release sites is unbiased.

Comparison with previous studies

Few methods have already been proposed to estimate both quantal and dynamic parameters from the same set of responses. Saviane and Silver (2006) applied multiple-probability fluctuation analysis to the first response in the train to obtain the quantal parameters. Using these estimates, they subsequently extracted the release probabilities (modified by short-term plasticity) and quantal sizes (reduced by postsynaptic receptor desensitization) for each of the subsequent responses in the train. In a third step, the time constant of facilitation was extracted from paired-pulse ratios. Finally, the time constant of the docking process was estimated by fitting the TM model to the release probabilities for the different responses in the train (while keeping τ_F fixed). Hallermann et al. (2010a,b) adopted a similar approach to compare the explanatory power of different models of repetitive transmission on the same dataset. Loebel et al. (2009) devised a somehow reversed procedure where, in a first step, least-squares fitting returns an estimate of the dynamic parameters, and of the product of the number of release sites and the quantal size (the so-called absolute synaptic efficacy). In the second step, another least-squares fitting returns the number of release sites (and thus the quantal size) that best reproduces the coefficient of variations of the synaptic responses.

All these methods proceed in a stepwise (and more or less laborious) manner using the estimates obtained in one step as fixed parameters in the next step. In this way, however, estimation errors in one step are propagated, and potentially amplified, in the next step. This should be of particular concern because the estimates in every step are obtained by least-squares fitting. All these methods are thus prone to the problems we have highlighted in the Maximum likelihood versus least-squares fit section, especially as the synaptic model is made more complex (i.e., features more parameters). Likewise, being based on least-squares fitting, all these methods require the repetition of the stimulation in identical conditions.

We are aware of only one other study (Costa et al., 2013) that has attempted to integrate the variability of the responses in the estimation of the synaptic parameters. There are, however, significant differences with our method. The method of Costa et al. (2013) does not take into account the correlation between different responses,

which, as we have shown, carries a significant amount of information. Furthermore, Gaussian noise is assumed in the underlying generative model, which prevents one from resolving the quantal parameters. Importantly, the variance of the Gaussian noise has to be estimated from the experimental responses, which requires one to repeat the same stimulation protocols. Costa et al. (2013) also report that least-squares fitting of TM-like models can result in grossly inaccurate estimates of the parameters. This is interpreted as a consequence of a shallow and rugged error surface, which can be flattened and deepened around the true minimum by using Poisson instead of regular spike trains (consistent with previous suggestions from the studies by Sen et al., 1996, and Varela et al., 1997, and with our results). In contrast with this interpretation, we have shown that for a certain combination of parameters and regular spike trains the least-squares fitting procedure is ill posed.

Advantages of the method

Our framework offers several advantages over existing state-of-the-art techniques for the characterization and quantification of synaptic transmission. The experimenter has complete freedom in choosing the stimulation protocol. For instance, the protocol can be chosen so as to elicit informative sequences of synaptic responses, thus achieving highly accurate estimates of the relevant parameters. As we have shown, for a given model, the asymptotic bounds on the accuracy of the estimates obtained with different protocols can be compared, before running any actual experiment, by computing the associated Fisher Information Matrices. Obviously, optimal protocols can be designed by using the same tool. Alternatively, and more interestingly, the stimulation protocol can be chosen so as to reproduce the statistical features of the *in vivo* spike trains driving the synaptic connections of interest. This would provide experimentalists as well as theoreticians with tools to develop effective descriptions of the transmission in physiologically relevant conditions. In fact, our method could already be fruitfully applied to investigate *in vivo* synaptic transmission at thalamocortical connections (Kaplan and Shapley, 1984; Ganmor et al., 2010).

To conclude, we would like to point out one potential application of our method that we deem especially interesting. Different mechanisms, both presynaptic and postsynaptic, have been proposed to be responsible for short-term synaptic plasticity (Zucker and Regehr, 2002; Fioravante and Regehr, 2011). It is presently unclear, however, whether these mechanisms are exclusive or, rather, they all cooperate to ensure the proper tuning of synaptic transmission across the wide range of possible patterns of presynaptic activity. Likelihood (once properly corrected for the differing number of free parameters) constitutes a theoretically principled metric to compare both the descriptive and predictive powers of different models over the same benchmark datasets. Our method, thus, paves the way to factorial testing of models of synaptic transmission, an approach that is increasingly and fruitfully being used in neuroscience and cognitive

science (Pinto et al., 2009; Daunizeau et al., 2011; van den Berg et al., 2014).

References

- Abbott L, Regehr WG (2004) Synaptic computation. *Nature* 431:796–803. [CrossRef](#) [Medline](#)
- Abbott LF, Varela JA, Sen K, Nelson SB (1997) Synaptic depression and cortical gain control. *Science* 275:220–224. [Medline](#)
- Barak O, Tsodyks M (2007) Persistent activity in neural network with dynamic synapses. *PLoS Comput Biol* 3:e25. [CrossRef](#)
- Bekkers J, Richerson G, Stevens C (1990) Origin of variability in quantal size in cultured hippocampal neurons and hippocampal slices. *Proc Natl Acad Sci U S A* 87:5359–5362. [Medline](#)
- Bertram R, Sherman A, Stanley EF (1996) Single-domain/bound calcium hypothesis of transmitter release and facilitation. *J Neurophysiol* 75:1919–1931. [Medline](#)
- Bhumbra GS, Beato M (2013) Reliable evaluation of the quantal determinants of synaptic efficacy using Bayesian analysis. *J Neurophysiol* 109:603–620. [CrossRef](#) [Medline](#)
- Blackman AV, Abrahamsson T, Costa RP, Lalanne T, Sjöström PJ (2013) Target-cell-specific short-term plasticity in local circuits. *Front Synaptic Neurosci* 5:11. [CrossRef](#) [Medline](#)
- Brémaud A, West DC, Thomson AM (2007) Binomial parameters differ across neocortical layers and with different classes of connections in adult rat and cat neocortex. *Proc Natl Acad Sci U S A* 104:14134–14139. [CrossRef](#) [Medline](#)
- Buzsáki G, Mizuseki K (2014) The log-dynamic brain: how skewed distributions affect network operations. *Nat Rev Neurosci* 15:264–278. [CrossRef](#) [Medline](#)
- Chabrol FP, Arenz A, Wiechert MT, Margrie TW, DiGregorio DA (2015) Synaptic diversity enables temporal coding of coincident multisensory inputs in single neurons. *Nat Neurosci* 18:718–727. [CrossRef](#) [Medline](#)
- Costa RP, Sjöström PJ, Van Rossum MC (2013) Probabilistic inference of short-term synaptic plasticity in neocortical microcircuits. *Front Comput Neurosci* 7:75. [CrossRef](#) [Medline](#)
- Daunizeau J, Preuschoff K, Friston K, Stephan K (2011) Optimizing experimental design for comparing models of brain function. *PLoS Comput Biol* 7:e1002280. [CrossRef](#) [Medline](#)
- Dayan P, Abbott LF (2001) *Theoretical neuroscience*. Cambridge, MA: MIT.
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Series B Stat Methodol* 39:1–38.
- Dittman JS, Kreitzer AC, Regehr WG (2000) Interplay between facilitation, depression, and residual calcium at three presynaptic terminals. *J Neurosci* 20:1374–1385.
- Dobrunz LE, Stevens CF (1997) Heterogeneity of release probability, facilitation, and depletion at central synapses. *Neuron* 18:995–1008. [Medline](#)
- Dobrunz LE, Stevens CF (1999) Response of hippocampal synapses to natural stimulation patterns. *Neuron* 22:157–166. [Medline](#)
- Efron B, Tibshirani RJ (1994) *An introduction to the bootstrap*. Boca Raton, FL: CRC.
- Fioravante D, Regehr WG (2011) Short-term forms of presynaptic plasticity. *Curr Opin Neurobiol* 21:269–274. [CrossRef](#) [Medline](#)
- Fortune ES, Rose GJ (2001) Short-term synaptic plasticity as a temporal filter. *Trends Neurosci* 24:381–385. [Medline](#)
- Fuhrmann G, Segev I, Markram H, Tsodyks M (2002) Coding of temporal information by activity-dependent synapses. *J Neurophysiol* 87:140–148. [Medline](#)
- Galassi M, Gough B, Jungman G, Theiler J, Davies J, Booth M, Rossi F (2009) *GNU scientific library: reference manual*. Bristol, U.K.: Network Theory Ltd.
- Ganmor E, Katz Y, Lampl I (2010) Intensity-dependent adaptation of cortical and thalamic neurons is controlled by brainstem circuits of the sensory pathway. *Neuron* 66:273–286. [CrossRef](#) [Medline](#)
- Gupta A, Wang Y, Markram H (2000) Organizing principles for a diversity of gabaergic interneurons and synapses in the neocortex. *Science* 287:273–278. [Medline](#)
- Hallermann S, Heckmann M, Kittel RJ (2010a) Mechanisms of short-term plasticity at neuromuscular active zones of drosophila. *HFSP J* 4:72–84. [CrossRef](#) [Medline](#)
- Hallermann S, Fejtova A, Schmidt H, Weyhersmüller A, Silver RA, Gundelfinger ED, Eilers J (2010b) Bassoon speeds vesicle reloading at a central excitatory synapse. *Neuron* 68:710–723. [CrossRef](#) [Medline](#)
- Hansel D, Mato G (2013) Short-term plasticity explains irregular persistent activity in working memory tasks. *J Neurosci* 33:133–149. [CrossRef](#) [Medline](#)
- Hardingham NR, Read JC, Trevelyan AJ, Nelson JC, Jack JJB, Bannister NJ (2010) Quantal analysis reveals a functional correlation between presynaptic and postsynaptic efficacy in excitatory connections from rat neocortex. *J Neurosci* 30:1441–1451. [CrossRef](#)
- Hempel CM, Hartman KH, Wang X-J, Turrigiano GG, Nelson SB (2000) Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex. *J Neurophysiol* 83:3031–3041. [Medline](#)
- Kandaswamy U, Deng P-Y, Stevens CF, Klyachko VA (2010) The role of presynaptic dynamics in processing of natural spike trains in hippocampal synapses. *J Neurosci* 30:15904–15914. [CrossRef](#) [Medline](#)
- Kaplan E, Shapley R (1984) The origin of the S (slow) potential in the mammalian lateral geniculate nucleus. *Exp Brain Res* 55:111–116. [Medline](#)
- Koester HJ, Johnston D (2005) Target cell-dependent normalization of transmitter release at neocortical synapses. *Science* 308:863–866. [CrossRef](#) [Medline](#)
- Larkman AU, Jack J, Stratford K (1997) Quantal analysis of excitatory synapses in rat hippocampal CA1 *in vitro* during low-frequency depression. *J Physiol* 505:457–471. [CrossRef](#)
- Larkman AU, Stratford K, Jack J (1991) Quantal analysis of excitatory synaptic action and depression in hippocampal slices. *Nature* 350:344–347. [CrossRef](#) [Medline](#)
- Loebel A, Silberberg G, Helbig D, Markram H, Tsodyks M, Richardson MJ (2009) Multiquantal release underlies the distribution of synaptic efficacies in the neocortex. *Front Comput Neurosci* 3:27. [CrossRef](#) [Medline](#)
- Loewenstein Y, Kuras A, Rumpel S (2011) Multiplicative dynamics underlie the emergence of the log-normal distribution of spine sizes in the neocortex *in vivo*. *J Neurosci* 31:9481–9488. [CrossRef](#) [Medline](#)
- Magleby K (1979) Facilitation, augmentation, and potentiation of transmitter release. *Prog Brain Res* 49:175–182. [CrossRef](#) [Medline](#)
- Markram H, Lübke J, Frotscher M, Roth A, Sakmann B (1997) Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J Physiol* 500:409–440. [CrossRef](#)
- Markram H, Wang Y, Tsodyks M (1998) Differential signaling via the same axon of neocortical pyramidal neurons. *Proc Natl Acad Sci U S A* 95:5323–5328. [Medline](#)
- Mongillo G, Barak O, Tsodyks M (2008) Synaptic theory of working memory. *Science* 319:1543–1546. [CrossRef](#) [Medline](#)
- Mongillo G, Hansel D, van Vreeswijk C (2012) Bistability and spatio-temporal irregularity in neuronal networks with nonlinear synaptic transmission. *Phys Rev Lett* 108:158101. [CrossRef](#) [Medline](#)
- Neher E, Sakaba T (2008) Multiple roles of calcium ions in the regulation of neurotransmitter release. *Neuron* 59:861–872. [CrossRef](#) [Medline](#)
- Oertner TG, Sabatini BL, Nimchinsky EA, Svoboda K (2002) Facilitation at single synapses probed with optical quantal analysis. *Nat Neurosci* 5:657–664. [CrossRef](#) [Medline](#)
- Oleskevich S, Clements J, Walmsley B (2000) Release probability modulates short-term plasticity at a rat giant terminal. *J Physiol* 524:513–523. [CrossRef](#)

- Pinto N, Doukhan D, DiCarlo JJ, Cox DD (2009) A high-throughput screening approach to discovering good forms of biologically inspired visual representation. *PLoS Comput Biol* 5:e1000579. [CrossRef](#) [Medline](#)
- Quastel D (1997) The binomial model in fluctuation analysis of quantal neurotransmitter release. *Biophys J* 72:728. [Medline](#)
- Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE* 77:257–286. [CrossRef](#)
- Richardson MJE, Silberberg G (2008) Measurement and analysis of postsynaptic potentials using a novel voltage-deconvolution method. *J Neurophysiol* 99:1020–1031. [CrossRef](#) [Medline](#)
- Rizzoli SO, Betz WJ (2005) Synaptic vesicle pools. *Nat Rev Neurosci* 6:57–69. [CrossRef](#) [Medline](#)
- Rothman JS, Cathala L, Steuber V, Silver RA (2009) Synaptic depression enables neuronal gain control. *Nature* 457:1015–1018. [CrossRef](#)
- Saviane C, Silver RA (2006) Fast vesicle reloading and a large pool sustain high bandwidth transmission at a central synapse. *Nature* 439:983–987. [CrossRef](#) [Medline](#)
- Scheuss V, Schneggenburger R, Neher E (2002) Separation of pre-synaptic and postsynaptic contributions to depression by covariance analysis of successive EPSCs at the calyx of held synapse. *J Neurosci* 22:728–739. [Medline](#)
- Sen K, Jorge-Rivera J, Marder E, Abbott L (1996) Decoding synapses. *J Neurosci* 16:6307–6318. [Medline](#)
- Senn W, Wyler K, Streit J, Larkum M, Lüscher H-R, Mey H, Müller L, Stainhauser D, Vogt K, Wannier T (1996) Dynamics of a random neural network with synaptic depression. *Neural Netw* 9:575–588. [CrossRef](#)
- Silver RA (2003) Estimation of nonuniform quantal parameters with multiple-probability fluctuation analysis: theory, application and limitations. *J Neurosci Methods* 130:127–141. [Medline](#)
- Silver RA, Lubke J, Sakmann B, Feldmeyer D (2003) High-probability unquantal transmission at excitatory synapses in barrel cortex. *Science* 302:1981–1984. [CrossRef](#) [Medline](#)
- Song S, Sjöström PJ, Reigl M, Nelson S, Chklovskii DB (2005) Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol* 3:e68. [CrossRef](#) [Medline](#)
- Stevens CF (2003) Neurotransmitter release at central synapses. *Neuron* 40:381–388. [Medline](#)
- Stone M (1974) Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc Series B Stat Methodol* 36:111–147.
- Trefethen LN, Bau D (1997) Numerical linear algebra, Vol 50. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Trigo FF, Sakaba T, Ogden D, Marty A (2012) Readily releasable pool of synaptic vesicles measured at single synaptic contacts. *Proc Natl Acad Sci U S A* 109:18138–18143. [CrossRef](#) [Medline](#)
- Tsodyks MV, Markram H (1997) The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc Natl Acad Sci U S A* 94:719–723. [Medline](#)
- Tsodyks MV, Uziel A, Markram H (2000) Synchrony generation in recurrent networks with frequency-dependent synapses. *J Neurosci* 20:1–5.
- van den Berg R, Awh E, Ma WJ (2014) Factorial comparison of working memory models. *Psychol Rev* 121:124–149. [CrossRef](#) [Medline](#)
- Varela JA, Sen K, Gibson J, Fost J, Abbott L, Nelson SB (1997) A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *J Neurosci* 17:7926–7940. [Medline](#)
- Wang Y, Markram H, Goodman PH, Berger TK, Ma J, Goldman-Rakic PS (2006) Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nat Neurosci* 9:534–542. [CrossRef](#) [Medline](#)
- Wu S, Wong KY, Tsodyks M (2013) Neural information processing with dynamical synapses. *Front Comput Neurosci* 7:188. [CrossRef](#) [Medline](#)
- Yuste R, Majewska A, Cash SS, Denk W (1999) Mechanisms of calcium influx into hippocampal spines: heterogeneity among spines, coincidence detection by NMDA receptors, and optical quantal analysis. *J Neurosci* 19:1976–1987. [Medline](#)
- Zucker RS, Regehr WG (2002) Short-term synaptic plasticity. *Annu Rev Physiol* 64:355–405. [CrossRef](#) [Medline](#)