UNIVERSITÄT BERN

Graduate School for Cellular and Biomedical Sciences

University of Bern

Statistical approaches for synaptic characterization

PhD Thesis submitted by

Camille Michel Jean-Claude Gontier

for the degree of

PhD in Neuroscience

Supervisor
Prof. Dr. Jean-Pascal Pfister
Department of Physiology
Faculty of Medicine of the University of Bern

Co-advisor
Prof. Dr. Martin Müller
Department of Molecular Life Sciences
University of Zurich

This work is licensed under a Creative Commons Attribution 4.0 International License
https://creativecommons.org/licenses/by/4.0/

source: https://doi.org/10.48549/3857 | downloaded: 27.4.2024

Accepted by the Faculty of Medicine, the Faculty of Science and the Vetsuisse Faculty of the University of Bern at the request of the Graduate School for Cellular and Biomedical Sciences

Bern, Dean of the Faculty of Medicine

Bern, Dean of the Faculty of Science

Bern, Dean of the Vetsuisse Faculty Bern

Acknowledgement

In 1159, English scholar John of Salisbury wrote in the *Metalogicon* that "*Bernard of Chartres used to compare us to dwarfs perched on the shoulders of giants. He pointed out that we see more and farther than our predecessors, not because we have keener vision or greater height, but because we are lifted up and borne aloft on their gigantic stature.*" Progress in science is made by standing on the shoulders of giants: *nanos gigantum humeris insidentes*. And some major giants have helped me during this doctoral work. As the American musician David Eric Grohl (almost) said, *there go my heroes*:

Jean-Pascal Pfister, for being the sword that helped me defeat the PhD monster[1], and for being the best PhD supervisor I ever had (sample size: $n = 1$). Jokes aside, thanks a lot for, in chronological order: giving me the chance to join again the wonderful world of academia after a two-year hiatus in the industry; for trusting me to join your team while I was barely able to derive the moments of a binomial distribution on the black board; for fostering a healthy and nourishing lab environment; for taking us snowshoeing in Kandersteg, hiking up to the Grünsee, and teaching us how to cook fondue in the middle of a frozen lake; for encouraging us to attend conferences from Lausanne to Barcelona, and from Lisbon to Princeton; and for making everything you could to make sure we would turn into successful and happy scientists.

Martin Müller, for accepting to be my co-advisor. I like to define myself as a theoretical researcher with a keen interest for collaborating with experimentalists and for bringing quantitative solutions to biological questions. This has been possible thanks to you: collaborating with you gave my purely theoretical work a different flavor.

Richard Naud, for accepting to serve as the external referee for this doctoral work, and for your insightful work on synaptic release [1].

Simone Carlo Surace, for the amazing job you did at correcting and optimizing my ~~terrible~~ legacy code, for the very insightful seminars on probability theory and information geometry, and for your technical and scientific knowledge and abilities, which were inspiring for all of us. I was very old years old when I learned that "KSE[2]" actually does not stand for "Kutschireiter-Surace Equation".

Anna Kutschireiter, for your kindness and your never-ending support, both in the lab and in the real world. This is all summarized by how you buoyed me up while I was terrified just before giving my very first lecture in front of students. After the lecture, I came back to a little note on my desk saying: "*See? They did not eat you.*

[1] `https://www.studyinternational.com/news/phd-students-mental-health/`

[2] `https://en.wikipedia.org/wiki/Kushner_equation`

Here, have a treat ☺", accompanied by a piece of Ragusa. To this date, this note is still religiously stuck to my copy of the Dayan and Abbott [2].

Christian "Chiseled" Horvat, for the vivid exchange of opinions, especially outside of the lab. Some of your ideas have been critically enlightening to me, and I will never be able to thank you enough for that.

Hui-An "Angela" Shen, for your constant cheerful disposition and curiosity. I'm still sorry we could not find a stuffed Olaf for your PhD defense: I used to believe that Stitch was cooler (because he knows how to pilot a spacecraft), but it turns out that Olaf has actually been on board a real spacecraft[3]!

Ehsan Abedi, for the decisive discussions about non-linear filtering on the white board, and for your impressive art[4]: to this date, your watercolor paintings are still decorating our office.

Jannes Jegminat, for the insightful discussions on science, philosophy, Life, the Universe, and Everything.

Gagan Narula, for the discussions on Hidden Markov Models, model selection, and on the Bayesian Information Criterion, and for your interactive presentations on zebra finches songs, putting back the "sound" into "sound scientific theories".

Igor Delvendahl, for sharing your data with us, for the fruitful collaboration between experiments and theory, and for your awesome salad.

Wayan Gauthey, for sharing your microscopy data with me and for carefully reviewing my `R` code. Good luck with the PhD in Princeton, Go Tigers!

Rui Ponte Costa, for the insightful seminar on neuronal data analysis, and for offering me to participate in your review on model-based approaches to synaptic characterization [3]: it meant a lot to the young scientist I was then.

Mihai "Mighty" Petrovici, for fostering an amazing lab environment, for the bouldering lessons, and for helping me move (twice). I just remembered that, during an especially high-level scientific argument (I can't remember whether it was about the best way to put coffee beans into the grinder, or the direction the toilet paper roll should be), you used an *argumentum ab auctoritate* and concluded that you should necessarily be right, since you have a PhD and I don't. Well, who's laughing now?

Jakob "Jacked" Jordan, for your interest in my idea of mixing arts and machine learning [4], for your thorough review of the draft and improvements in data analysis: without you and your help in this side project, this thesis would have been completed way earlier.

Walter Senn, for the support and exchange of ideas.

Henrik "Hardcore" Mettler, for bravely accepting to go fly with me. I am sorry I nearly killed you, I swear my landing skills got better now.

And all members of the Computational Neuroscience Group and Neuro-inspired theory, modeling and applications (TMA) Group: **Federico "Furious" Benitez,**

[3] More precisely in the Soyuz TMA-15M launched on November 23rd 2014, as seen in `https://www.youtube.com/watch?v=ZfUf2uwGhyk`

[4] `https://www.facebook.com/UniBern/videos/a-portrait-of-herman-hesse-in-100-hours/3557047710999043/`

Laura "Lovely" Kriener, Matthias "Masculine" Tsai, Paul "Poetic" Haider, Nicolas "Na(ugh)tty" Deperrois, Elena Kreutzer, Kevin "Kinky" Max, Katharina "Kamikaze" Wilmes, Benjamin "Bootylicious" Ellenberger, Arno Joeffry Granier, Ben "Badass" von Hünerbein, Virginie "Voracious" Sabado, Timo "Tumultuous" Gierlich, Simon "Salacious" Benjamin Brandt, and Lukas Sebastian Huber[5].

[5]Don't ask me where these nicknames are coming from. Camille "Curvy" Gontier is the first one to find them dumb.

Abstract

Synapses are fascinatingly complex transmission units. One of the fundamental features of synaptic transmission is its stochasticity, as neurotransmitter release exhibits variability and possible failures. It is also quantised: postsynaptic responses to presynaptic stimulations are built up of several and similar quanta of current, each of them arising from the release of one presynaptic vesicle. Moreover, they are dynamic transmission units, as their activity depends on the history of previous spikes and stimulations, a phenomenon known as synaptic plasticity. Finally, synapses exhibit a very broad range of dynamics, features, and connection strengths, depending on neuromodulators concentration [5], the age of the subject [6], their localization in the CNS or in the PNS, or the type of neurons [7].

Addressing the complexity of synaptic transmission is a relevant problem for both biologists and theoretical neuroscientists. From a biological perspective, a finer understanding of transmission mechanisms would allow to study possibly synapse-related diseases, or to determine the locus of plasticity and homeostasis. From a theoretical perspective, different normative explanations for synaptic stochasticity have been proposed, including its possible role in uncertainty encoding, energy-efficient computation, or generalization while learning. A precise description of synaptic transmission will be critical for the validation of these theories and for understanding the functional relevance of this probabilistic and dynamical release.

A central issue, which is common to all these areas of research, is the problem of synaptic characterization. Synaptic characterization (also called synaptic interrogation [8]) refers to a set of methods for exploring synaptic functions, inferring the value of synaptic parameters, and assessing features such as plasticity and modes of release. This doctoral work sits at the crossroads of experimental and theoretical neuroscience: its main aim is to develop statistical tools and methods to improve synaptic characterization, and hence to bring quantitative solutions to biological questions.

In this thesis, we focus on model-based approaches to quantify synaptic transmission, for which different methods are reviewed in Chapter 3. By fitting a generative model of postsynaptic currents to experimental data, it is possible to infer the value of the synapse's parameters. By performing model selection, we can compare different modelizations of a synapse and thus quantify its features. The main goal of this thesis is thus to develop theoretical and statistical tools to improve the efficiency of both model fitting and model selection.

A first question that often arises when recording synaptic currents is how to precisely observe and measure a quantal transmission. As mentioned above, synaptic transmission has been observed to be quantised. Indeed, the opening of a single presynaptic vesicle (and the release of the neurotransmitters it contains) will create a stereotypical postsynaptic current q, which is called the quantal amplitude. As the number of

activated presynaptic vesicles increases, the total postsynaptic current will increase in step-like increments of amplitude q. Hence, at chemical synapses, the postsynaptic responses to presynaptic stimulations are built up of k quanta of current, where k is a random variable corresponding to the number of open vesicles. Excitatory postsynaptic current (EPSC) thus follows a multimodal distribution, where each component has its mean located to a multiple kq with $k \in \mathbb{N}$ and has a width corresponding to the recording noise σ. If σ is large with respect to q, these components will fuse into a unimodal distribution, impeding the possibility to identify quantal transmission and to compute q. How to characterize the regime of parameters in which quantal transmission can be identified? This question led us to define a practical identifiability criterion for statistical model, which is presented in Chapter 4. In doing so, we also derive a mean-field approach for fast likelihood computation (Appendix A) and discuss the possibility to use the Bayesian Information Criterion (a classically used model selection criterion) with correlated observations (Appendix B).

A second question that is especially relevant for experimentalists is how to optimally stimulate the presynaptic cell in order to maximize the informativeness of the recordings. The parameters of a chemical synapse (namely, the number of presynaptic vesicles N, their release probability p, the quantal amplitude q, the short-term depression time constant τ_D, etc.) cannot be measured directly, but can be estimated from the synapse's postsynaptic responses to evoked stimuli. However, these estimates critically depend on the stimulation protocol being used. For instance, if inter-spike intervals are too large, no short-term plasticity will appear in the recordings; conversely, a too high stimulation frequency will lead to a depletion of the presynaptic vesicles and to a poor informativeness of the postsynaptic currents. How to perform Optimal Experiment Design (OED) for synaptic characterization? We developed an Efficient Sampling-Based Bayesian Active Learning (ESB-BAL) framework, which is efficient enough to be used in real-time biological experiments (Chapter 5), and propose a link between our proposed definition of practical identifiability and Optimal Experiment Design for model selection (Chapter 6).

Finally, a third biological question to which we ought to bring a theoretical answer is how to make sense of the observed organization of synaptic proteins. Microscopy observations have shown that presynaptic release sites and postsynaptic receptors are organized in ring-like patterns, which are disrupted upon genetic mutations. In Chapter 7, we propose a normative approach to this protein organization, and suggest that it might optimize a certain biological cost function (e.g. the mean current or SNR after vesicle release).

The different theoretical tools and methods developed in this thesis are general enough to be applicable not only to synaptic characterization, but also to different experimental settings and systems studied in physiology. Overall, we expect to democratize and simplify the use of quantitative and normative approaches in biology, thus reducing the cost of experimentation in physiology, and paving the way to more systematic and automated experimental designs.

Contents

List of Figures

List of Tables

Acronyms

ACh Acetylcholine. 2

AD Alzheimer's Disease. 11

AIC Akaike Information Criterion. 15

AP Action Potential. 2

AZ Active Zone. 2

BAL Bayesian Active Learning. 22

BF Bayes Factor. 15

BIC Bayesian Information Criterion. 15

BRP Bruchpilot. 2

CNS Central Nervous System. ix

EM Expectation-Maximization. 13

EPSC Excitatory postsynaptic current. x

EPSP Excitatory postsynaptic potential. 4

FIM Fisher Information Matrix. 20

GABA γ-Aminobutyric Acid. 2

GLM Generalized Linear Models. 23

GluR Glutamate Receptor. 2

GPP Gibbs Point Process. 101

gSTED Time-Gated Stimulated Emission Depletion. 2

HMM Hidden Markov Model. 6

i.i.d. Independent and Identically Distributed. 13

ISI Inter-Spike Interval. 5

MAP Maximum A Posteriori. 32

MCMC Markov Chain Monte Carlo. 22

mEPSC Miniature Excitatory postsynaptic current. 9

MH Metropolis-Hastings. 13

MLE Maximum Likelihood Estimator. 15

NMJ Neuro-Muscular Junction. 2

NPF Nested Particle Filter. 73

OED Optimal Experiment Design. x

PD Parkinson's Disease. 11

PNS Peripheral Nervous System. ix

PPR Paired-pulsed ratio. 7

RMSE Root-Mean-Square Error. 76

SAMs Synaptic Adhesion Molecules. 2

SNR Signal-to-Noise Ratio. x

STB Short-Term Biphasic Plasticity. 7

STD Short-Term Depression. 5

STDP Spike Timing-Dependent Plasticity. 11

STF Short-Term Facilitation. 5

STP Short-Term Plasticity. 4

VGCC Voltage-Gated Calcium Channels. 2

To my family and Isa, who helped me a lot during this work; and to Miti, who did not help me at all, but who is very cute.

Chapter 1

Introduction

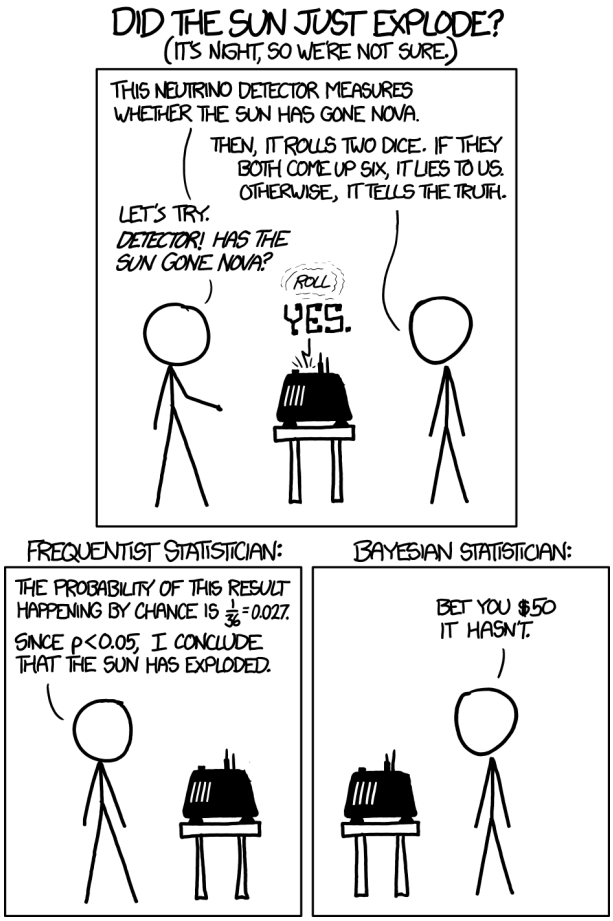


Figure 1.1: On the importance of using the correct statistical tools[1].

1.1 The biological features of synaptic transmission

Synapses are fascinatingly complex transmission units. Across species and parts of the nervous system, they exhibit a very broad range of dynamics, features, and connection strengths, depending on neuromodulators concentration [5], the age of the

[1] `https://xkcd.com/1132/`, use with authorization.

subject [6], their localization in the CNS or in the PNS, or the type of neurons [7]. Despite this diversity, all chemical synapses share the same basic principles and structures, illustrated in Figure 1.2. The presynaptic Active Zone (AZ) and the postsynaptic density are separated by the narrow synaptic cleft, and are stabilized close to each other by Synaptic Adhesion Molecules (SAMs). Synaptic vesicles contain neurotransmitter molecules, such as glutamate (the most common excitatory neurotransmitter in vertebrates), GABA (inhibitory), or ACh (at the Neuro-Muscular Junction or NMJ). Upon the arrival of an Action Potential (AP) in the presynaptic terminal, Voltage-Gated Calcium Channels (VGCC) will increase the calcium concentration in the presynaptic terminal, leading to the fusion of vesicles (from a pool of readily-releasable vesicles) with the plasma membrane and to the release of neurotransmitters in the cleft. Neurotransmitter molecules will then bind with postsynaptic receptors, which will either create an immediate flux of ions (for ionotropic receptors) or initiate an indirect chemical pathway (for metabotropic receptors); in both cases, it will result in an excitatory or inhibitory postsynaptic current.

At a finer level, the spatial organization of synaptic proteins can be measured using either confocal or gSTED microscopy (Figure 1.3). At the Drosophila NMJ, Bruchpilot (BRP) presynaptic proteins are located close to vesicle release sites, around the VGCC at the center of the AZ, while postsynaptic GluR (glutamate receptors) are located at the postsynaptic density.

In most cases, the presynaptic terminal is located on the axon of a neuron, while the postsynaptic bouton is located on a dendrite or on the soma of another neuron. However, different arrangements also exist: axon-to-axon, dendrite-to-dendrite, axon-to-bloodstream, soma-to-dendrite, dendrite-to-soma, or soma-to-soma. Especially, at the NMJ, chemical signal is passed to a target muscle cell. Finally, one should also note the existence of electrical synapses, in which the pre- and postsynaptic cells are directly connected by gap junctions. They allow for a faster and direct transmission of currents (without intermediary chemical transmitters), but cannot modify the gain of the signal.

One of the fundamental features of synaptic transmission is its **stochasticity**, as neurotransmitter release exhibits variability and possible failures: upon presynaptic stimulation, only a fraction of the vesicles in the readily-releasable pool will indeed fuse with the synaptic membrane and release their neurotransmitters. This fraction is highly variable, both when considering successive stimulations of the same synapse (trial-to-trial, or within-subject, variability) or when comparing different synapses (between-subject variability). For a given synapse, the mean fraction averaged over different trials corresponds to the **release probability** p. Incidentally, single vesicles may also randomly release their neurotransmitters even in the absence of presynaptic spike. This stochastic release of a single vesicle is known as **spontaneous release**, as opposed to **evoked release** where several vesicles fuse simultaneously following a presynaptic spike.

Synaptic transmission at chemical synapses is also **discrete**: the spontaneous release of one vesicle will create a small, although detectable, postsynaptic current, called a **miniature postsynaptic current (mPSC)**, or "mini". It has been observed that, following evoked release, the postsynaptic current was always a multiple of the mPSC. Hence, in chemical synapses, the postsynaptic responses to presynaptic stimulations are built up of several and similar quanta of current, each of them arising from the

release of one presynaptic vesicle[2] [9]. The current elicited by the release of one presynaptic vesicle is called the **quantal amplitude** q.

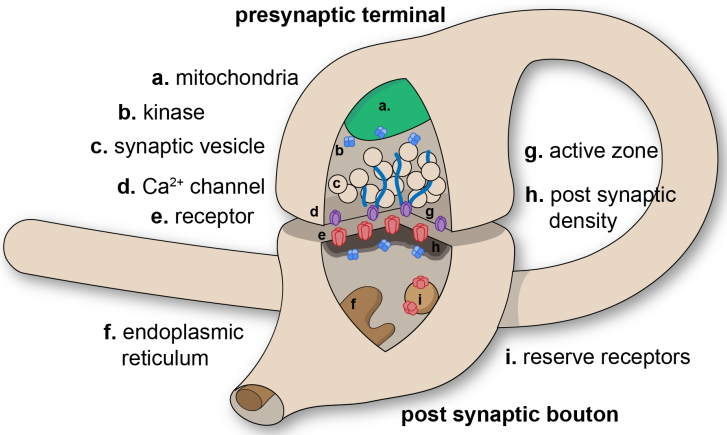


Figure 1.2: **Diagram of a chemical synaptic connection**. This diagram shows a stereotypical chemical synapse, where a chemical signal is passed from one neuron (*presynaptic terminal*) to another (*postsynaptic bouton*). In most cases, the presynaptic terminal is located on the axon of a neuron, while the postsynaptic bouton is located on a dendrite or on the soma of another neuron: however, different arrangements also exist. Especially, at the NMJ, chemical signal is passed to a target muscle cell. The presynaptic Active Zone (AZ) and the postsynaptic density are separated by the narrow synaptic cleft, and are stabilized close to each other by Synaptic Adhesion Molecules (SAMs). Synaptic vesicles contain neurotransmitter molecules, such as glutamate (the most common excitatory neurotransmitter in vertebrates), GABA (inhibitory), or ACh (at the NMJ). Upon the arrival of an action potential in the presynaptic terminal, Voltage Gated Calcium Channels (VGCC) will increase the calcium concentration in the presynaptic terminal, leading to the fusion of vesicles with the plasma membrane and to the release of neurotransmitters in the cleft. Neurotransmitter molecules will then bind with postsynaptic receptors, and trigger either an excitatory (depolarizing) or inhibitory (hyperpolarizing) postsynaptic current. Image from [10] (CC BY-SA 4.0 license).

A classical biophysical model used to describe this stochastic release of neurotransmitter is called the **binomial model**: upon the arrival of an action potential in the presynaptic terminal, vesicles are released with a given probability p. The binomial model [12] assumes that there are N independent release sites and that for each site the release probability p is identical. Therefore, the number of released vesicles after spike i, k_i, is distributed according to a binomial distribution with parameters N and p. This model further assumes that each vesicle release gives rise to a quantal current q, such that the overall Excitatory postsynaptic current (EPSC) is given by $e_i = qk_i + \epsilon_i$, where ϵ_i models measurement noise[3] typically drawn from a standard normal distribution with standard deviation σ. The distribution of outputs is given

[2]The quantal nature of synaptic transmission had been identified in the 1950s by Bernard Katz, who later received the Nobel Prize in medicine for his research.

[3]σ accounts for different sources of variability, including the thermal noise of the amplifier used for the recording, or the natural fluctuations of the membrane cell which are not accounted for in the model.

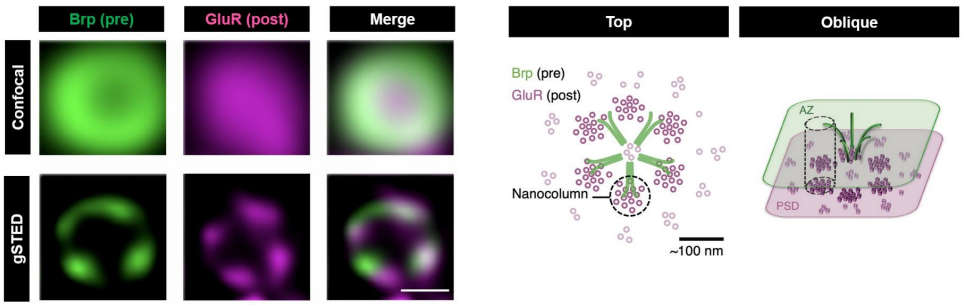


Figure 1.3: **Spatial organization of synaptic proteins**. At the Drosophila NMJ, synaptic proteins' position can be measured using either confocal or gSTED microscopy. Bruchpilot (BRP) presynaptic proteins are located close to vesicle release sites, around the VGCC at the center of the AZ. Postsynaptic GluR (glutamate receptors) are located at the postsynaptic density. Courtesy of Wayan Gauthey and Martin Müller, University of Zurich [11].

by

$$p(e_i) = \sum_{k_i=0}^{N} p(e_i|k_i)p(k_i) \tag{1.1}$$

where $[e_i|k_i] \sim \mathcal{N}(qk_i, \sigma^2)$ and $k_i \sim \text{Bin}(N, p)$.

Under this binomial model, a synapse is thus described by a vector Θ of 4 parameters: the number of vesicles N, the release probability p, the quantal amplitude q [mA], and the recording noise σ [mA] (Figure 1.4). In most cases, these parameters cannot be measured directly, but can be inferred using EPSCs[4] recorded on the postsynaptic side and elicited by stimulating the presynaptic cell, a method known as **synaptic characterization**. Different methods to infer the values of the parameters of a synapse will be described in Section 1.5.

It is important to note that the binomial model detailed above and having parameters $\Theta = [N, p, q, \sigma]$ is not the only one that can be used to fit synaptic data: synaptic transmission can be explained using a wide range of possible models, of different complexities and levels of plausibility, modelling different features and having different parameters. Different models of the synapse will be presented in Section 1.3.

1.2 Computational models of synaptic Short-Term Plasticity

Synapses have been observed to be dynamic transmission units, as their activity depends on the history of previous spikes and stimulations. In this section, we focus on a particularly important feature of synaptic transmission, which is called Short-Term Plasticity (STP), and which is illustrated in Figure 1.5. In the basic binomial

[4]In practice, the binomial model can be applied to both EPSPs and EPSCs. In any postsynaptic cell, the current (EPSC) causes a change in membrane voltage (EPSP), which also depends on the membrane properties (i.e. membrane time constant and resistance). Up to a certain point, these factors are linear and thus EPSP can be used for quantal analysis as well. One of the major differences is that postsynaptic currents have faster kinetics than the corresponding voltage changes, which allows e.g. responses to high-frequency repetitive stimulation to be resolved. In addition, voltage changes (EPSPs) often do not summate linearly when the voltage changes are large (they can also elicit spiking of the postsynaptic cell). Clamping voltage and recording the currents circumvents this.

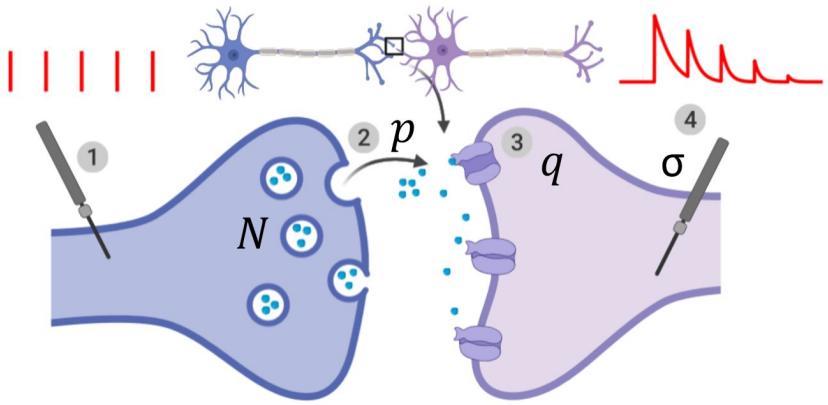


Figure 1.4: **First illustration of the binomial model (without STP)**. (1): the presynaptic button is artificially stimulated. Red vertical bars show 5 presynaptic spikes with a constant interspike interval. (2): these evoked stimuli lead to neurotransmitter release. After spike i, k_i vesicles (out of the N vesicles) release their neurotransmitter with a probability p. (3): neurotransmitters bind to receptors and elicit a postsynaptic current. A single release event triggers a quantal response of amplitude q. (4): the recorded postsynaptic response after spike i is the sum of the effects of the k_i release events. EPSCs correspond to the amplitude of each peak of the postsynaptic response to a presynaptic spike.

model, the refilling dynamics of vesicles is neglected, implying that all recordings $(e_i)_{1\leq i\leq T}$ are mutually independent. However, in practice, synaptic transmission is dynamic and exhibits plasticity:

- **Short-Term Depression (STD)** corresponds to a progressive decrease in the amplitude of successive postsynaptic currents upon repeated activation of the presynaptic cell [2]. The mechanism mainly responsible for STD is the depletion of the pool of presynaptic vesicles: after fusing with the cell membrane and releasing their neurotransmitters, these vesicles and their content are retrieved by endocytosis and recycled to be reused for the next stimulation. STD depends on the rate at which vesicles refill during Inter-Spike Intervals (ISI). If the stimulation frequency is higher than the rate of refilling dynamics, the pool of readily-releasable vesicles will progressively be depleted, leading to less neurotransmitters being released at each stimulation and hence to smaller postsynaptic currents (see [13] for a detailed review on synaptic machinery). Recording from a depressing synapse can be seen in Figure 1.6.

- Inversely, **Short-Term Facilitation (STF)** corresponds to a progressive increase in the amplitude of successive postsynaptic currents: the more the presynaptic cell is stimulated, the higher the postsynaptic current. STF is due to an increase in calcium concentration in the presynaptic cell after each spike generation, which is going to increase the vesicles' release probability.

Different computational models of short-term synaptic plasticity have been proposed, with different levels of complexity and representing different features of synaptic dynamics (see [14] and [15] for a detailed review). The idea that the readily-releasable vesicles represent a fraction $x(t)$ of the total pool of vesicles, which is depleted upon presynaptic stimulation and exponentially decays back to 1, was first formulated by Liley and North [16]. The evolution of the fraction of available vesicles can be repre-

sented by a differential equation:

$$\frac{dx(t)}{dt} = \frac{1 - x(t)}{\tau_D} - p \cdot x(t^-) \cdot \sum_i \delta(t - t_i) \tag{1.2}$$

where $x(t)$ is the fraction of filled vesicles among a total pool of N, and τ_D is their replenishment time constant. The first term in the right-hand side corresponds to the natural replenishment of vesicles with time constant τ_D, while the second term corresponds to depletion at each spike time $(t_i)_{1 \leq i \leq T}$. The dynamic quantity $x(t)$ is evaluated at t^-, i.e. immediately before the spike [15]. However, this equation alone can account for depression, but not for facilitation. A second differential equation can be added:

$$\frac{dp(t)}{dt} = \frac{p_0 - p(t)}{\tau_F} + p_0 \cdot (1 - p(t^-)) \cdot \sum_i \delta(t - t_i) \tag{1.3}$$

where $p(t)$ is the release probability (which is not a constant anymore but evolves depending on the history of synaptic activity), and p_0 its resting value (i.e. the release probability at the beginning of the stimulation). Once again, the first term in the right-hand side corresponds to an exponential decrease of the release probability to its resting value p_0, while the second term corresponds to an increase at each spike.

The **Tsodyks-Markram model** [17] corresponds to the set of two equations:

$$\frac{dx(t)}{dt} = \frac{1 - x(t)}{\tau_D} - p(t^-) \cdot x(t^-) \cdot \sum_i \delta(t - t_i)$$

$$\frac{dp(t)}{dt} = \frac{p_0 - p(t)}{\tau_F} + p_0 \cdot (1 - p(t^-)) \cdot \sum_i \delta(t - t_i)$$

A similar system of equations is used in [18] and [19], adding a third differential equation:

$$\frac{dv(t)}{dt} = \frac{v_0 - v(t)}{\tau_m} + J \cdot p(t^-) \cdot x(t^-) \cdot \sum_i \delta(t - t_i)$$

where $v(t)$ is the postsynaptic membrane potential, v_0 its resting value, τ_m the membrane constant of the postsynaptic cell, and J a scalar synaptic efficacy parameter.

The Tsodyks-Markram model is a non-stochastic, mean-field, approximation of synaptic transmission. The mean of the EPSC at time t can be computed as $\langle e_t \rangle = N \cdot q \cdot p(t) \cdot x(t)$. However, such a deterministic approach to short-term plasticity only allows to model averages, and neglects correlations between successive postsynaptic responses. In [5] and [20], a discrete-time and stochastic variation of the Tsodyks-Markram model is used: the synapse is modelled as a Hidden Markov Model (HMM) with hidden states n_i (the number of vesicles in the filled state at the moment of spike i) and k_i (the number of vesicles among the n_i which will release the neurotransmitters following spike i).

Between two successive spikes, the probability that a given vesicle is refilled within a time Δ_{t_i} is $I_i(\Delta_{t_i}) = 1 - \exp\left(-\frac{\Delta_{t_i}}{\tau_D}\right)$, which corresponds to the first term in the right-hand side of Eq. 1.2. The stochastic release of neurotransmitters (which corresponds to the second term in the right-hand side of Eq. 1.2) is represented by the number of open vesicles k_i being drawn from a binomial distribution with parameters n_i and p_i. Similarly, the release probability p_i can be computed recursively as $p_i = p_0 + p_{i-1}(1 - p_0)\exp\left(-\frac{\Delta_{t_i}}{\tau_F}\right)$.

By setting the values of the parameters τ_D, τ_F, and p_0, the Tsodyks-Markram model allows to reproduce depressing, facilitating, or stable synapses. However, it cannot explain more complex dynamics; for instance, it has been observed that vesicles refilling rate may not be constant, but might increase upon sustained presynaptic stimulation [15]. Although the exact mechanisms for this enhanced vesicle replenishment are still unknown, it is hypothesized that vesicle refilling dynamics depend on the intracellular calcium concentration, which increases upon sustained presynaptic stimulation. Two different models have been proposed to account for enhanced vesicle replenishment. In [21], the authors propose that the refilling time constant τ_D in Eq. 1.2 is actually time dependent, a phenomenon called Frequency-Dependent Recovery (FDR):

$$\frac{d\tau_D(t)}{dt} = \frac{\tau_{D_0} - \tau_D(t)}{\tau_{\text{FDR}}} - a_{\text{FDR}} \cdot \tau_D(t^-) \cdot \sum_i \delta(t - t_i)$$

In the absence of stimulation, $\tau_D(t)$ decays back to its resting value τ_{D_0} with a time constant τ_{FDR}, while each spike reduces it by $a_{\text{FDR}} \cdot \tau_D(t^-)$ (with $0 < a_{\text{FDR}} < 1$).

Furthermore, it has been proposed that the baseline release probability p_0 in Equation 1.3 might itself depend on presynaptic activity, and be governed by a similar differential equation:

$$\frac{dp_0(t)}{dt} = \frac{\tilde{p}_0 - p_0(t)}{\tau_j} - a_j \cdot p_0(t^-) \cdot \sum_i \delta(t - t_i)$$

This activity-dependent decrease in release probability, which may contribute to STD, might be related to VGCC inactivation [15].

Finally, another dynamical feature that is worth mentioning is Short-Term Biphasic Plasticity (STB). STB is a combination of depression and facilitation, in which the Paired-pulsed ratio (PPR) first increases and then decreases: it has been observed, for instance, in some hippocampal interneurons [22]. Dittman et al. [23] proposed a model in which the replenishment rate is also a function of the residual presynaptic calcium, which concentration follows a first order differential equation with an exponentially decaying term and a jump term depending on the presynaptic spikes times.

1.3 Different models of the synapse

The standard binomial model introduced in Section 1.1 relies on several assumptions and simplifications. Firstly, it neglects vesicle refilling and calcium uptake dynamics, and posits that different postsynaptic currents are independent and do not depend on the history of previous spikes and activations. Different models of Short-Term

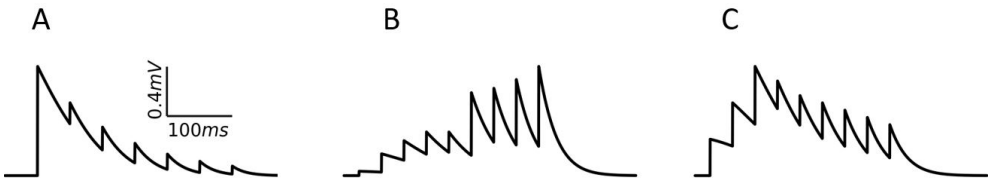


Figure 1.5: **Illustration of Short-Term Synaptic Plasticity**. (a) Example of depression of an excitatory synapse, as seen between two layer 5 pyramidal neurons recorded from a rat somatosensory cortex. (b) Example of facilitation of an excitatory synapse, as seen between a pyramidal cell and an inhibitory interneuron in layer 2/3 from a rat somatosensory cortex. Examples redrawn from [2]. (c) Example of biphasic plasticity, as observed in hippocampal interneurons. Example redrawn from [14].

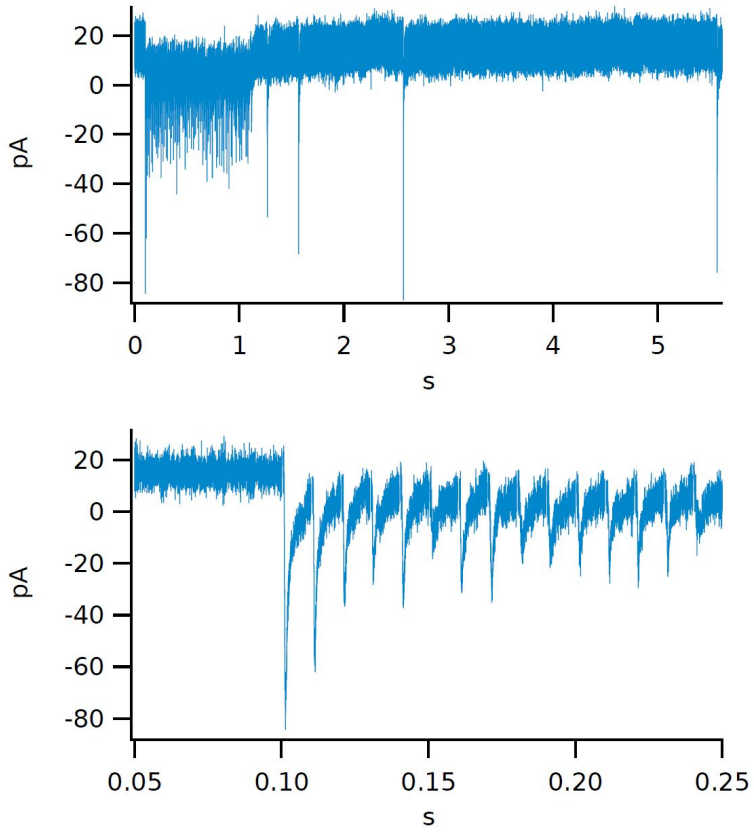


Figure 1.6: **Recording of EPSCs from a depressing synapse**. Postsynaptic current recorded from a cerebellar mossy fiber to granule cell synapse. The presynaptic cell is stimulated using a repetitive train stimulation protocol consisting of 100 stimuli at 100 Hz followed by 6 recovery pulses at increasing intervals (upper plot). Recovery from depression can be seen in the recovery pulses as EPSCs amplitudes get larger with longer ISIs. Conversely, depression can be seen during the high-frequency tetanic stimulation phase (lower plot) as the vesicles pool gets depleted. Courtesy of Igor Delvendahl and Martin Müller, University of Zurich.

Plasticity have been described in the previous Section 1.2. Secondly, it assumes that the quantal amplitudes q are homogeneous and identical for each release site. It further posits that all release sites are similar and have the same homogeneous refilling time constant τ_D. Finally, it assumes that the total postsynaptic current

will be the sum of the effect of independent vesicle releases and neglects non-linear phenomena such as postsynaptic receptors saturation.

Starting from the classical binomial model with parameters $\Theta = [N, p, q, \sigma]$, one can thus construct different models of increasing complexities by relaxing some of these assumptions. Added features to a statistical model can be thought of as added free parameters.

Short-term Plasticity

A first assumption of the classical binomial model is that it neglects vesicle refilling and calcium uptake dynamics, and posits that different postsynaptic currents are independent and do not depend on the history of previous spikes and activations. However, as described in the previous section, synaptic transmission has been observed to be dynamic, with short-term depression depending on the refilling of presynaptic vesicles between successive presynaptic stimulations, and short-term facilitation depending on the calcium uptake at each spike.

Depression and facilitation can thus be accounted for, and added as features to the model, by adding the extra free parameters τ_D and τ_F.

Quantal amplitude variability

A second assumption of the classical binomial model is that the quantal amplitudes q are homogeneous and identical for each release site. However, mEPSC amplitudes have been observed not to be exactly similar and stereotypical, but rather to follow a unimodal distribution [1]. The effect of one release event q can thus be thought of as a random variable instead of a constant.

Assuming that quantal amplitudes are normally distributed, EPSC e_i will be computed as

$$e_i = \sum_{j=1}^{k_i} q_j + \epsilon_i$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ is the recording noise, $k_i \sim \text{Bin}(N, p)$ the number of released vesicles, and the quantal amplitude q_j follows a normal distribution with mean μ_q and standard deviation σ_q.

Another possibility is to represent the radius of presynaptic vesicles as a random variable: assuming that vesicles are approximately spherical [1], the quantal amplitude can be computed as being proportional to the vesicle's volume, i.e. $q_j = \alpha_q \frac{4}{3}\pi r_j^3$ with r_j being drawn e.g. from a normal distribution[5]: $r_j \sim \mathcal{N}(\mu_V, \sigma_V^2)$.

Vesicles pool heterogeneity

A third assumption of the classical binomial model is that release sites are similar and have the same homogeneous refilling time constant τ_D. However, it has been suggested [24] that presynaptic vesicles actually belong to two different groups: a readily releasable pool and a reluctantly releasable pool, having different refilling time scales. Assuming that the fast-refilling pool and the slow-refilling pool have respective refilling time constants $\tau_D^{(1)}$ and $\tau_D^{(2)}$, and that the fast-refilling pool makes up a fraction A of the total number of vesicles, then the recovery (i.e. the evolution

[5]Which std. σ_V is assumed to be sufficiently small for the radius r_j not to be negative.

of the fraction of available vesicles $x(t)$ after presynaptic depletion) can be fitted with a biexponential trajectory [25]:

$$x(t) = A(1 - e^{-t/\tau_D^{(1)}}) + (1 - A)(1 - e^{-t/\tau_D^{(2)}})$$

It is to be observed that this model will also have a higher number of hidden states. Instead of having two hidden variables n_i (corresponding to the number of available vesicles before spike i) and k_i (corresponding to the number of released vesicles), there will be 4 hidden variables corresponding to each subpool $n_i^{(1)}$, $n_i^{(2)}$ and $k_i^{(1)}$, $k_i^{(2)}$.

Postsynaptic saturation

Finally, the classical binomial model posits that the total postsynaptic current will be the sum of the effects of independent vesicle releases. In practice, neurons often display non-linear behaviors, such as spike-frequency adaptation [26, 27] or postsynaptic saturation[6] [29].

Saturation of postsynaptic receptors is believed to play a major role at depressing synapses, and can be modelled as follows [15]: assuming that $D(t)$ is a scalar variable indicating the fraction of postsynaptic receptors in a non-desensitized state, it will evolve as

$$\frac{dD(t)}{dt} = \frac{1 - D(t)}{\tau_s} - a_s \cdot p(t^-) \cdot x(t^-) \cdot D(t^-) \cdot \sum_i \delta(t - t_i)$$

where τ_s is the time constant of recovery from desensitization. The mean of the EPSC at time t would thus be computed as $\langle e_t \rangle = N \cdot q \cdot p(t) \cdot x(t) \cdot D(t)$ (see Section 1.2).

This list is not exhaustive, as many models can be proposed. So far, we have seen that the basic properties of synaptic transmission (that is to say the stochastic and quantised release of neurotransmitters) can be represented by a simple binomial model having parameters $\Theta = [N, p, q, \sigma]$; and that the values of these parameters can be inferred by fitting the model on observed EPSCs. Other features of synaptic transmission (plasticity, quantal variability, vesicles heterogeneity, and postsynaptic saturation) can be modelled by extending the classical binomial model, thus describing other models with more free parameters. As we will see in the next section, synaptic characterization is thus not only about inferring the values of the parameters for a given model (Section 1.5), but also about finding the right model to fit the data (Section 1.6).

1.4 What is synaptic characterization about ?

Addressing the complexity of synaptic transmission, i.e. being able to derive an accurate model of the studied synapse, is a relevant problem for both biologists and theoretical neuroscientists.

Firstly, from a theoretical point of view, modifications of synaptic transmission properties are believed to underlie learning, memory and, more generally, neural dynamics:

[6] Although it has been observed not to be an issue for some synapses, e.g. at the Calyx of Held [28].

to better understand how the brain works, it is thus necessary to precisely quantify how its fundamental brick (i.e. the synapse) is working. Especially, it has been speculated that synaptic variability and unreliability could be more of a feature than a bug and could serve a computational purpose. In [30], the authors show that "*unreliable synapses serve to increase cortical firing by means of increasing membrane fluctuations, similar to high conductance states*". Moreover, from a network perspective, stochastic synaptic transmission could be similar to drop-out in artificial neural networks: it could be beneficial at the computational level by helping the network to generalise better and being more energy efficient [31]. Different normative explanations for synaptic stochasticity have been proposed, including its possible role in uncertainty encoding, energy-efficient computation, or generalization while learning (Figure 1.7). A precise description of synaptic transmission will be critical for the validation of these theories and for understanding the functional relevance of this probabilistic and dynamical release.

Secondly, being able to precisely infer the values of synaptic parameters is critical for studying synaptic plasticity and homeostasis [32, 33]. For instance, when studying Spike Timing-Dependent Plasticity (STDP), comparing the values of N, p, and q before and after the stimulation protocol allows to determine the locus of plasticity [34, 35, 36].

Finally, from a biological perspective, obtaining reliable estimates of synaptic parameters, and more generally a finer understanding of transmission mechanisms, is a required step to study possibly synapse-related diseases, such as schizophrenia, Parkinson's Disease (PD), Alzheimer's Disease (AD), and addiction [37]. Synaptic density has been shown to be an important correlate of cognitive decline [38]: an interesting area of research would be to correlate decline with the characteristics of individual synapses.

A central issue, which is common to all these areas of research, is the problem of synaptic characterization. Synaptic characterization (also called synaptic interrogation [8]) refers to a set of methods for exploring synaptic functions, inferring the value of synaptic parameters, and assessing features such as plasticity and modes of release. Here, we focus on model-based approaches to quantify synaptic transmission, for which different methods are reviewed in Chapter 3. Synaptic characterization is thus about:

- **Parameters inference**: by fitting a generative model of postsynaptic currents to experimental data, it is possible to infer the value of the synapse's parameters. Different methods for model fitting will be reviewed in Section 1.5 and in Chapter 3, but existing methods all suffer from different drawbacks: they might only be applicable to specific synapses, be biased or lead to inaccurate estimates of the parameters, or require long computation times.

- **Model selection**: by determining which model (among a set of possible candidate models) is the best one to fit experimental data, we can detect and quantify the features exhibited by a synapse. Although a critical step in statistical analysis, model selection is rarely performed in synaptic characterization, and the question of model identifiability (which we discuss in Section 1.7 and Chapter 4) is often disregarded.

Overall, our main goal is to develop theoretical and statistical tools to improve the efficiency of both model fitting and model selection.

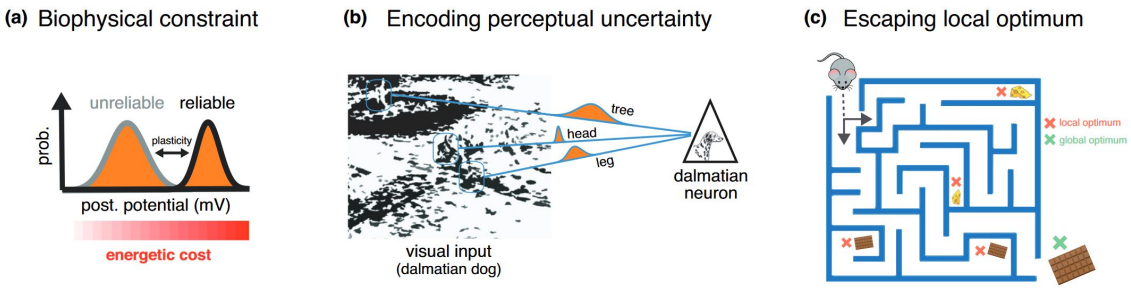


Figure 1.7: **Possible computational roles of synaptic stochasticity.** (a) Reliable synaptic transmission implies a high energetic cost (represented by the red colour bar). Such biophysical constraints may explain stochasticity at the synapse level, as limited energy supply may only allow reliable synapses to develop if necessary. (b) Stochastic synaptic transmission may allow the brain to encode sensory statistics, and especially perceptual uncertainty. To infer a high-level information (e.g. the presence of a dalmatian dog), different lower-level contextual information need to be integrated. The uncertainty of the connections representing different features should be proportional to how relevant or how certain they are. (c) By adding noise to the system, stochastic synapses have been suggested to enable neural networks to generalize better, to avoid overfitting and to escape local optima. Figure from [39], reproduced with authorisation.

1.5 Methods for synaptic parameter inference

From the simple binomial model explained above, it follows (neglecting STP and recording noise), that the distribution of EPSCs has a mean $\mu = Npq$ and a variance $\sigma^2 = q^2Np(1-p)$. The release probability p directly depends on the calcium concentration of the solution (increasing $[Ca^{2+}]$ leading to a higher p), which can be modified experimentally. Since the mean μ varies linearly with p, and since $\sigma^2 = \mu q - \frac{\mu^2}{N}$, plotting the variance as a function of the mean will result in a parabolic plot. This plot can be fitted to observations obtained under various calcium concentrations in order to infer the values of N and p. This widely used method is called **mean-variance analysis** [40, 41, 42, 43]. Its main drawback is that it requires to perform recordings under different calcium concentrations, and, for each of these settings, to obtain enough data points to accurately estimate the mean and the variance.

The **cumulative EPSC method** [44] consists in using a continuous high-frequency stimulation of the presynaptic cell, to completely deplete the pool of readily-releasable vesicles and reach steady state depression. Cumulative EPSC amplitudes are then plotted as a function of time: fitting the last points of the plot by linear regression allows to estimate the rate of vesicles replenishment (i.e. the slope of the fitted line) while back-propagating the fitted line to time 0 gives an estimate of Nq (see Figure 4(C) in [44]). However, this method is only valid for depressing synapses and also requires inter-stimulation intervals to be constant. Different methods based on high-frequency activations are discussed in details in [45].

In the **method of moments**, the moments of the distribution of EPSCs are expressed as a function of the synaptic parameters, and are then matched to the observation's moments. Besides the aforementioned mean $\mu = Npq$ and variance $\sigma^2 = q^2Np(1-p)$, it is also possible to approximate the maximum EPSC as Nq and the failure rate as $(1-p)^N$. In [46], the authors further use Fourier analysis to obtain an estimate of

the quantal amplitude q. As in mean-variance analysis, this method requires a very large number of observations to correctly estimate the variance.

Finally, **density estimation** methods consist in optimizing the parameters of a probability distribution so as to match the observed distributions of EPSCs. In [47], the authors fit the distribution of postsynaptic currents using mixtures of unimodal distributions. The goodness-of-fit for different numbers of components is assessed using the Wilks statistic to infer synaptic parameters. In more robust **model-based approaches**, the following ingredients are used:

- A generative model $\mathcal{M}$, with parameters θ.
- A set of T observations $\mathcal{D} = \{e_i\}_{1 \leq i \leq T}$.
- The corresponding ISIs $\Delta_{1:T}$, i.e. the times at which the presynaptic cell has been stimulated to record $\mathcal{D}$.

Once these elements have been specified, it becomes possible to compute explicitly the likelihood function $\mathcal{L}(\theta) = p(\mathcal{D}|\theta, \Delta_{1:T}, \mathcal{M})$, i.e. the likelihood of the observations given the stimulation times and the parameters.

The likelihood function $\mathcal{L}$ can then be used:

- in different optimization techniques, such as the Expectation-Maximization (EM) algorithm [20] or genetic optimization [48], to obtain a point-based estimate of the parameters $\hat{\theta} = \arg\max_\theta \mathcal{L}(\theta)$ (Figure D.1);
- in the Metropolis-Hastings (MH) algorithm to compute the full posterior distribution of parameters $p(\theta|\mathcal{D}, \Delta_{1:T}, \mathcal{M}) \propto \mathcal{L}(\theta)p(\theta|\mathcal{M})$ [5] (Figure D.2), where $p(\theta|\mathcal{M})$ is the prior distribution of the parameters of model $\mathcal{M}$;
- to update the weights of the particles in particle filtering schemes [49] (Figure D.3).

These model-based methods are reviewed in Chapter 3. A major step, which is common to all model-based approaches (whether their aim is to obtain a simple point-based estimate of the parameters or their complete posterior distribution), is the computation of the likelihood function $\mathcal{L}(\theta)$. For models in which observations in $\mathcal{D}$ are Independent and Identically Distributed (i.i.d.), e.g. the binomial model without short-term plasticity, the likelihood can be easily computed[7] as $p(\mathcal{D}|\theta, \mathcal{M}) = \prod_i p_\theta(e_i)$, where the computation time of $p_\theta(e_i)$ scales linearly with N (see Eq. 1.1).

However, many models of synaptic transmission (especially models of STP) assume that successive observations are correlated. In the general case, the likelihood is computed as

$$p(\mathcal{D}|\theta, \Delta_{1:T}, \mathcal{M}) = p(e_1) \prod_{i=2}^{T} p_\theta(e_i|e_{1:i-1})$$

More specifically, depressing and facilitating synapses can be modelled as a Hidden Markov Model, with hidden states n_i (the number of vesicles in the filled state at the moment of spike i) and k_i (the number of vesicles among the n_i which will release

[7] For simplicity, we will drop non-ambiguous notations and denote $p_\theta(\cdot) = p(\cdot|\theta, \Delta_{1:T}, \mathcal{M})$.

the neurotransmitters following spike i) and observation e_i. The state transition probability is defined as

$$p_\theta(n_i, k_i|n_{i-1}, k_{i-1}) = p_\theta(k_i|n_i)p_\theta(n_i|n_{i-1}, k_{i-1})$$

where $p_\theta(k_i|n_i) \sim \text{Bin}(n_i, p_i)$ and the refilling probability $p_\theta(n_i|n_{i-1}, k_{i-1})$ depends on Δ_{t_i} and τ_D; and the emission probability is $p_\theta(e_i|n_i, k_i) = p_\theta(e_i|k_i) \sim \mathcal{N}(qk_i, \sigma^2)$.

The joint distribution of the observations $\mathcal{D}$ and hidden states $\mathbf{n} = n_{1:T}$ and $\mathbf{k} = k_{1:T}$ is thus computed as

$$p_\theta(\mathcal{D}, \mathbf{n}, \mathbf{k}) = p_\theta(e_1|k_1)p_\theta(k_1|n_1)p_\theta(n_1)\prod_{i=2}^{T} p_\theta(e_i|k_i)p_\theta(k_i|n_i)p_\theta(n_i|n_{i-1}, k_{i-1})$$

The Baum-Welch algorithm is an exact forward-backward procedure which can be used to compute the likelihood $\mathcal{L}(\theta) = \sum_{\mathbf{n},\mathbf{k}} p_\theta(\mathcal{D}, \mathbf{n}, \mathbf{k})$. However, its algorithmic complexity will scale with N^4, making it unpractical for studying synapses with a large number of vesicles or for on-the-loop applications. This is a serious impediment to synaptic characterization, since likelihood computation is at the basis of all model-based approaches. In Chapter 5, we derive a particle filtering scheme, which highly parallelizable structure enables fast computations for online applications. Besides, in Appendix A we derive mean-field approximations to reduce the algorithmic complexity of likelihood computation to N^2.

1.6 Theory of model selection

In the previous section (1.5), we presented different methods allowing, for a given model $\mathcal{M}$ and a set of observations $\mathcal{D}$, to infer the values of the parameters of $\mathcal{M}$. However, synaptic characterization is not only about estimating the parameters of a model, but also about finding out which model should be fitted on the observations in the first place: how to pick the right $\mathcal{M}$ to fit and describe the studied synapse? **Model selection** is the process by which a model is chosen out of a pre-defined set of possible candidates (see [50] for a detailed review).

Models are compared based on their ability to explain and fit the data, while being penalized for their complexity (usually, for their number of free parameters) to avoid overfitting (Figure 1.8). The latter objective corresponds to regularization: its goal is to prune the model, and to penalize the likelihood to prevent overfitting. Finding the correct model (or, at least, the one amidst the family of models we have at hand that is the best suited to fit our data; as George Box said, *"all models are wrong, but some are useful"* [8] [52]) is thus a complex[9] and crucial first step in accurately inferring synaptic parameters.

Different **model selection criteria** have been defined:

[8] As stated by William Bialek in [51], referring to another famous paper: *"The American philosopher Lawrence Peter (Yogi) Berra is reported to have quipped about a restaurant that 'nobody goes there anymore, it is too crowded'. Perhaps some papers are so famous that nobody reads them anymore."*

[9] Which is nicely illustrated by Norbert Wiener's aphorism: *"[T]he best material model of a cat is another, or preferably the same cat."*[53]

- The **model evidence** for a model $\mathcal{M}$ given data $\mathcal{D}$ is simply defined as

$$p(\mathcal{D}|\mathcal{M}) = \int d\theta p(\mathcal{D}|\theta, \mathcal{M}) p(\theta|\mathcal{M})$$

where $p(\theta|\mathcal{M})$ is the prior distribution of parameters θ for model $\mathcal{M}$. Two competing models $\mathcal{M}$ and $\mathcal{M}'$ can be compared using the **Bayes Factor (BF)**:

$$\frac{p(\mathcal{D}|\mathcal{M})}{p(\mathcal{D}|\mathcal{M}')} = \frac{\int d\theta p(\mathcal{D}|\theta, \mathcal{M}) p(\theta|\mathcal{M})}{\int d\theta p(\mathcal{D}|\theta, \mathcal{M}') p(\theta|\mathcal{M}')}$$

Depending on the value of this ratio, one can determine if the evidence for $\mathcal{M}$ compared to $\mathcal{M}'$ is negative, not significant, or decisive, and thus conclude as to which model is the best one to fit the data $\mathcal{D}$. An interesting reference on how it differs from frequentist hypothesis testing is [54]. However, in practice the evidence $p(\mathcal{D}|\mathcal{M})$ is often intractable for complex models, as it requires to integrate marginals for each parameter.

- An interesting approximation of the model evidence is given by the **Bayesian Information Criterion (BIC)** [55]:

$$\mathrm{BIC}_{\mathcal{M}}(\mathcal{D}) := -2 \log \mathrm{p}(\mathcal{D}|\widehat{\theta}, \mathcal{M}) + \mathrm{k}_{\mathcal{M}} \log(\mathrm{T}) \approx -2 \log \mathrm{p}(\mathcal{D}|\mathcal{M})$$

where $\hat{\theta} = \arg\max_\theta p(\mathcal{D}|\theta, \mathcal{M})$ is the Maximum Likelihood Estimator (MLE) of θ, $k_{\mathcal{M}} = \dim(\Theta)$ is the number of independent parameters of $\mathcal{M}$, and T is the number of data points in $\mathcal{D}$. In the limit of large T, the BIC provides a reliable approximation of the model evidence (see Section 4.6.1 for a detailed derivation): the lower the BIC of model $\mathcal{M}$, the better its score. The BIC is the sum of two terms: a likelihood term $-2 \log p(\mathcal{D}|\hat{\theta}, \mathcal{M})$ which represents the ability of the model $\mathcal{M}$ to explain $\mathcal{D}$, and a penalty term $k_{\mathcal{M}} \log(T)$ which favors simpler models and prevents overfitting, as explained in Section 4.3.5.

- Another widely used selection criterion is the **Akaike Information Criterion (AIC)** [56]:

$$\mathrm{AIC}_{\mathcal{M}}(\mathcal{D}) := -2 \log \mathrm{p}(\mathcal{D}|\widehat{\theta}, \mathcal{M}) + 2\mathrm{k}_{\mathcal{M}} \approx -2\mathbb{E}_{\mathcal{D}'}(\log \mathrm{p}(\mathcal{D}'|\widehat{\theta}(\mathcal{D}), \mathcal{M}))$$

Compared to the BIC, the AIC penalizes the number of free parameters in the models differently. It is derived as an unbiased estimator of the Kullback-Leibler divergence between the respective distributions of observations according to the ground-truth model and to $\mathcal{M}$ (see [50] for a detailed computation). In [57], the authors discuss cases in which the AIC is more suited than the BIC to compare models, i.e. is more likely to correctly select the ground-truth model: in the tapering-effects context (i.e. when the family of competing models shows a continuous increase of complexity, as illustrated in Fig. 1 in [57]), the AIC will perform better than the BIC (as the BIC-selected model is likely to underfit, especially at low T). But when candidate models only have a few important features (Fig. 2 in [57]), the BIC should be favored, while the AIC-selected model is likely to overfit[10].

It is important to note that both the BIC and AIC were originally derived assuming that elements in $\mathcal{D}$ are i.i.d., i.e. that $p(\mathcal{D}) = p(e_{1:T}) = \prod_i p(e_i)$.

[10] See also `https://stats.stackexchange.com/a/493214/271601`

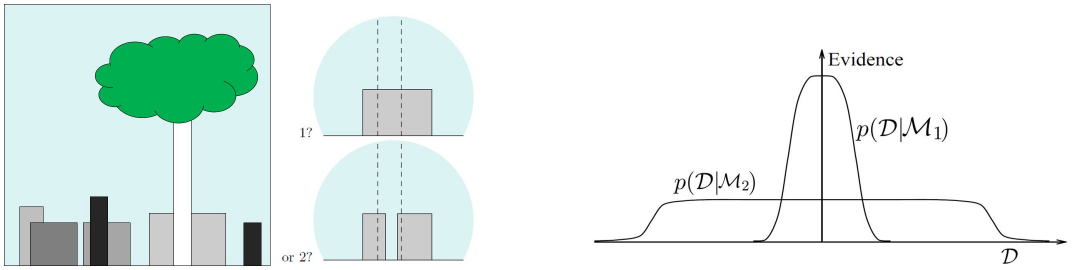
This is a very restrictive assumption, as many models of synaptic transmission (especially models of STP) assume that successive observations are correlated, i.e. that $p(e_i, e_{i+1}) = p(e_{i+1}|e_i)p(e_i) \neq p(e_{i+1})p(e_i)$. The implications of this assumption, and the possibility to use model selection criteria for correlated data, are discussed in Chapter 4 and in Appendix B.

- Other model selection criteria, having different theoretical justifications, include the Widely Applicable Information Criterion (WAIC) [58], the Widely Applicable Bayesian Information Criterion (WBIC) [58], the Focused Information Criterion (FIC) [59], or the Hannan-Quinn Information criterion (HQC) [60].
- **Cross-validation** [61] is mostly reserved to situations in which the number of data points T is sufficiently large to be split into a training and a test set without impairing training (e.g. in machine learning training). In experimental settings, where the number of available observations can be small, overfitting is avoided by penalizing the likelihood of the model with an ad hoc regularization term (which is generally dependent on the number of free parameters in the model, see above). Interestingly, cross-validation has been shown to be equivalent to the AIC in the limit of large T [62, 59], while leave-v-out cross validation is equivalent to the BIC for $v = T(1 - \frac{1}{\log T - 1})$ [63].

Bayesian model selection is widely used for performing model comparison and hypothesis testing, and has been applied to a broad range of domains: to study the volume of synaptic vesicles [1], group studies [64], astronomy [65], and exoplanets [66], to name a few. Other fields related to Bayesian model selection are Bayesian Model Averaging [67] (in which the prediction is the weighted sum of the predictions of the models, weighted by their respective evidences), or Optimal Experiment Design (see Chapter 5).

Although a critical first step prior to parameter inference, model selection is often disregarded in synaptic characterization studies [20, 5]. Moreover, while performing model selection, two main pitfalls need to be avoided:

- **Overfitting**, i.e. choosing a generative model $\mathcal{M}$ that is more complicated than the ground-truth studied synapse, which would lead to poor generalization and prediction, and inaccurate parameters inference. Model selection criteria precisely penalize model complexity to prevent overfitting.
- **Underfitting**, i.e. choosing a generative model $\mathcal{M}$ that is simpler than the ground-truth studied synapse. Indeed, some features of the studied synapse (e.g. quantal release or short-term plasticity) may not be experimentally observable, depending on the experimental protocol being used. Large recording noise, or poorly chosen stimulation protocols, could make interesting features of the studied system undetectable and may not allow to draw conclusions on the nature of the synapse. For instance, when observing that a Gaussian distribution provides a better fit to the observations than a binomial release model, one might be tempted to conclude that the synapse indeed releases only one quantum at a time ("uni quantal hypothesis"). However, it is also possible that the synapse is actually binomial ("multi quantal hypothesis"), but that quantal peaks on the histogram of recorded EPSCs are hidden due to noisy or scarce data points. Similarly, when no Short-Term Plasticity appears in postsynaptic currents, this can be due either to the synapse being intrinsically static, or to plasticity being hidden due poorly chosen stimulations times. When a simpler model provides a better fit (in the sense of model selection) than a more



(a) (b)

Figure 1.8: **Model Comparison and Occam's Razor**. (a) How many boxes are located behind the tree: 1 (model $\mathcal{M}_1$) or 2 (model $\mathcal{M}_2$)? Both models explain the observation equally well. However, $\mathcal{M}_1$ is simpler, and $\mathcal{M}_2$ should be discarded under Occam's Razor. Indeed, it has more free parameters, and implies that two boxes having exactly the same height and color are behind the tree. (b) Why Bayesian inference encompasses Occam's razor, and why complex models should be discarded if a simpler one explains the same data. The x-axis represents all the possible data observations $\mathcal{D}$. The predictions of a model $\mathcal{M}$ are quantified by a probability distribution $p(\mathcal{D}|\mathcal{M})$, which is normalized over $\mathcal{D}$: $\int d\mathcal{D} p(\mathcal{D}|\mathcal{M}) = 1$. Models differ in the range of $\mathcal{D}$ they can explain: a model with more free parameters will explain a larger range of observations. If data can be well explained by the simpler model $\mathcal{M}_1$, its evidence will be larger than that of $\mathcal{M}_2$ due to the normalization. Figures redrawn from [68].

complicated hypothetical release model, it is impossible to conclude whether it is due to the intrinsic nature of the synapse or to data not being sufficiently informative, an issue linked to the question of **identifiability**.

1.7 Identifiability

A question that may arise when recording synaptic currents is how to precisely observe and measure a quantal transmission. As mentioned above, synaptic transmission has been observed to be quantised. Indeed, the opening of a single presynaptic vesicle (and the release of the neurotransmitters it contains) will create a stereotypical postsynaptic current q, which is called the quantal amplitude. As the number of activated presynaptic vesicles increases, the total postsynaptic current will increase in step-like increments of amplitude q. Hence, in chemical synapses, the postsynaptic responses to presynaptic stimulations are built up of k quanta of current, where k is a random variable corresponding to the number of open vesicles. EPSCs thus follow a multimodal distribution, where each component has its mean located to a multiple kq with $k \in \mathbb{N}$ and has a width corresponding to the recording noise σ (Figure 1.9, A). If σ is large with respect to q, these components will fuse into a unimodal distribution, impeding the possibility to identify quantal transmission and to compute q (Figure 1.9, B).

A similar question may arise when studying synaptic short-term plasticity. Indeed, if the release probability p is low, or if inter-spike intervals are large compared to plasticity time constants τ_D and τ_F, then no sign of short-term plasticity will appear in postsynaptic recordings (Figure 1.10). In both cases, due to the intrinsic values of the ground-truth parameters and of the chosen inputs to the system, results of model selection would be inconsistent with the ground-truth model. In the first case, the

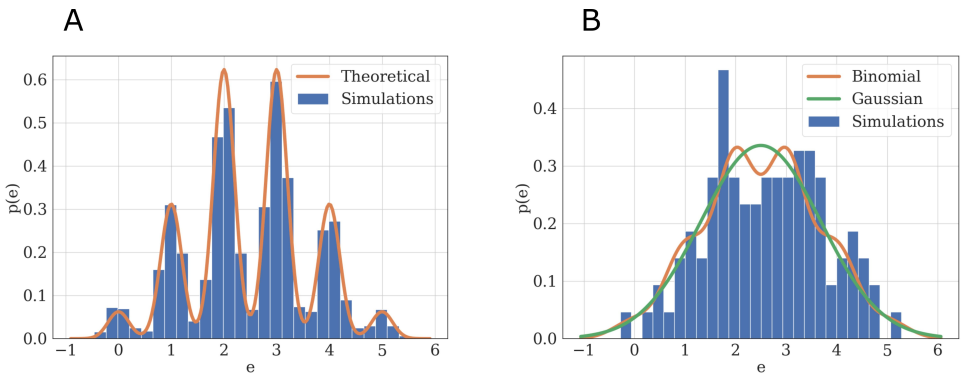


Figure 1.9: **Quantal and stochastic transmitter release.** (A) Blue: histogram of 2,000 simulated EPSCs generated from the binomial model with parameters $N = 5$, $p = 0.5$, $q = 1$, $\sigma = 0.2$. Orange: theoretical distribution of EPSCs from Eq. 1.1. Due to step-like quantal fluctuations in EPSC amplitude [69], the distribution appears multimodal and can be described by a mixture of Gaussian functions. The width of each mode corresponds to the recording noise σ, while the height of the k-th mode corresponds to $p(k)$. (B) Illustration of practical non-identifiability. Blue: histogram of 100 simulated EPSCs with $N = 5$, $p = 0.5$, $q = 1$, $\sigma = 0.4$ (i.e. same parameters as in A, but with a smaller number of observations and higher σ). Due to the small number of data points and high recording noise σ, the binomial parameters can only be loosely estimated, which is characterized by the fact that a Gaussian distribution (green) will provide a better fit to the data than a binomial distribution (orange).

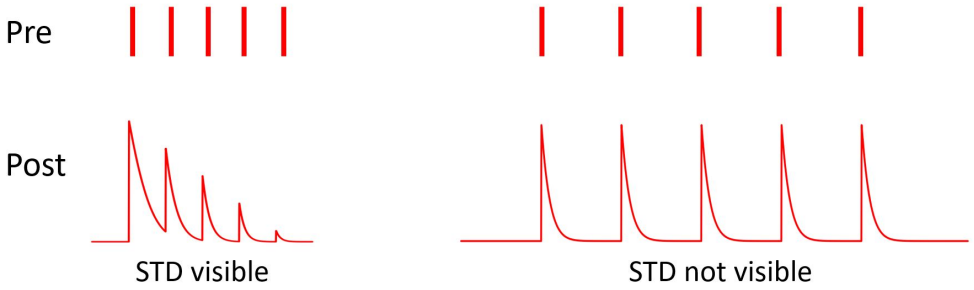


Figure 1.10: **Non-identifiability of Short-Term Depression.** In these surrogate data, an artificial depressing synapse is being studied. The presynaptic axon is stimulated using evoked action potentials with constant ISIs (upper plots), while the amplitudes of postsynaptic responses (lower plots) are being recorded. On the left, the ISIs between successive presynaptic spikes are sufficiently short (i.e. shorter than the depression time constant τ_D), and STD can be seen in the decreasing current amplitudes upon successive stimulations (lower left). On the right, for the same synapse, choosing ISIs that are longer than τ_D makes STD undetectable (lower right).

ground truth model is a binomial model with a multimodal distribution of EPSCs, but due to the high σ model selection would favor a simpler unimodal distribution. In the second case, the ground truth model is a model with short-term depression, but due to the chosen stimulation protocol model selection would favor a simpler static model.

Synaptic characterization, and more generally parameter inference in any field, require that the model $\mathcal{M}$ used is identifiable. Two notions need to be defined: **structural** identifiability and **practical** identifiability.

1.7.1 Structural identifiability

Structural (i.e. model-based) identifiability is a property of the model, regardless of experimental results. In a structurally identifiable system, the dimension of the output is sufficiently high with respect to the dimension of the parameters vector to uniquely define it: the parameters can be non-ambiguously inferred if the complete distribution of the output is known. Structural identifiability has been widely studied in many fields of physics and biology [70, 71, 72, 73, 74], and different criteria exist to assess the structural identifiability of a model [75].

When applied to our system, the structural identifiability of the binomial model is equivalent to the distribution of outputs $p_\theta(\mathcal{D})$ being uniquely defined by θ. Consistently with [70, 75], we can define the structural identifiability domain Θ_S of a model $\mathcal{M}$ with parameter space Θ as:

$$\Theta_S = \{\theta \in \Theta \mid (\forall \theta' \in \Theta)\ \theta \neq \theta' \iff p(\mathcal{D}|\theta, \mathcal{M}) \neq p(\mathcal{D}|\theta', \mathcal{M})\} \tag{1.4}$$

Similarly, $\mathcal{M}$ is said to be structurally identifiable if $\Theta = \Theta_S$. Intuitively, if θ is in the structural identifiability domain of $\mathcal{M}$, it can be uniquely identified from $p(\mathcal{D}|\theta, \mathcal{M})$.

1.7.2 Practical identifiability

Structural identifiability is a theoretical property and is not equivalent to practical (i.e. experiment-based) identifiability, which is a property of both the model and the experimental protocol: a model which is structurally identifiable might lead to a poor practical identifiability of parameters if data points are noisy or scarce. The accuracy of model-based methods for inferring the values of parameters depends on the experimental protocol used to record the data, as observations need to be sufficiently informative to allow a correct estimation of the parameters. Contrary to structural identifiability (see criterion 1.4), a quantitative criterion is lacking for practical identifiability, which is usually only qualitatively assessed. Non-practical identifiability refers to regimes in which parameters can only be loosely estimated; but one would need to define what does "loose" mean.

A definition for the practical identifiability of a parameter has previously been proposed in [70], along with an approach for detecting practical non-identifiabilities based on the profile likelihood [76, 77]. The authors first define the likelihood-based confidence intervals for the estimator $\hat{\theta}_i$ of the i-th parameter of a model $\mathcal{M}$:

$$C_{i,\Delta} = \{\theta_i \mid \mathcal{L}(\hat{\theta}_i|\mathcal{D}) - \mathcal{L}(\theta_i|\mathcal{D}) < \Delta\}$$

where

$$\mathcal{L}(\theta_i|\mathcal{D}) = \max_{\theta_{j\neq i}} \mathcal{L}(\theta|\mathcal{D})$$

for a given threshold Δ. Then, they propose the following definition: *A parameter estimate $\hat{\theta}_i$ is practically non-identifiable, if the likelihood-based confidence region is infinitely extended in increasing and/or decreasing direction of θ_i, although the likelihood has a unique minimum for this parameter*, meaning that the decrease in

likelihood compared to the optimal parameters estimate stays below the threshold Δ in direction of θ_i.

A first limitation of this definition is to be data-dependent: it only holds for a specific set of recorded data $\mathcal{D}$. Indeed, likelihood-based confidence intervals, and hence practical identifiability, are defined with respect to a certain data set $\mathcal{D}$, and may thus vary for different realizations of the experiment. A second limitation is that it requires an arbitrary criterion Δ on the possible precision of parameters estimate.

Other statistical tools can also be used to assess identifiability. For $\theta = [\theta_1, \theta_2, \dots, \theta_M]$, the Fisher Information Matrix (FIM) is a $M \times M$ matrix defined as follows:

$$[\mathcal{I}(\theta)]_{i,j} = \mathbb{E}_{p_\theta(\mathcal{D})}\left((\frac{\partial}{\partial \theta_i} \log p_\theta(\mathcal{D})) \cdot (\frac{\partial}{\partial \theta_j} \log p_\theta(\mathcal{D})) \right)$$

where $\mathbb{E}_{p_\theta(\mathcal{D})}(\cdot)$ denote the expectation over the responses sequences $\mathcal{D}$.

The FIM can be used to evaluate the identifiability of a system [71]. A set of parameters θ is said to be locally identifiable if no other set of parameters in its neighborhood can generate the same distribution $p_\theta(\mathcal{D})$. This is equivalent to $\mathcal{I}(\theta)$ being invertible: the number of identifiable parameters is thus equal to the number of non-zero eigenvalues of $\mathcal{I}(\theta)$.

For a finite number of data points T, a lower bound on the normalized standard deviation of the estimate of the parameter θ_i can be obtained *a priori* using the Cramer-Rao bound:

$$\epsilon(\theta_i) \geq \frac{\sqrt{[\mathcal{I}(\theta)]^{-1}_{ii}}}{\sqrt{T}\theta_i} \tag{1.5}$$

where $\mathcal{I}(\theta)$ is the Fisher Information Matrix of the distribution p_θ [2]. A criterion for practical identifiability could be based on the expected variance of the estimator, but would once again require to set an arbitrary criterion on the said variance.

In Chapter 4, we propose a non-arbitrary definition of practical identifiability, by transforming a model identifiability problem into a model selection problem: a model is said to be practically identifiable when its parameters can be correctly inferred given a certain experimental protocol. But, as explained previously, different possible models can be fitted on a data set. Recorded data need to be sufficiently informative not only to give a correct estimate of the parameters of a model, but also to select the correct model (i.e. the model from which they have been generated). We argue that a model is practically identifiable if and only if it is also correctly identified as the model providing the best fit to the data. For a given experimental protocol, we define the practical identifiability domain of a statistical model as the set of parameters for which the model is correctly identified as the ground truth compared to a simpler alternative submodel.

1.8 Optimal Experiment Design

Optimizing the parameters of an experiment is a critical problem in general physiology. Experimental design usually involves setting a plethora of parameters: which stimulations or perturbations to perform, which dosage or concentrations to use, how to determine sample size or measurement time points, etc. These parameters will

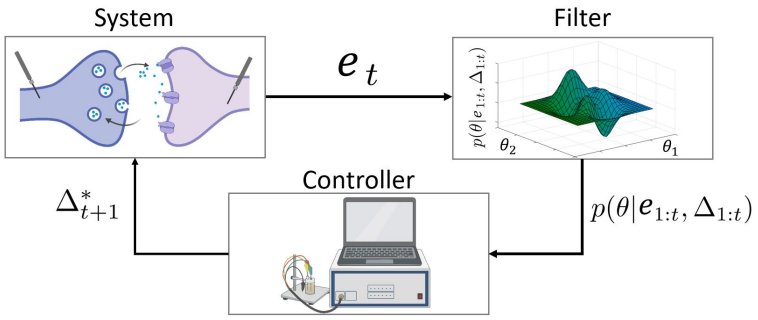


Figure 1.11: **Optimal *infomax* design of experiment using Bayesian Active Learning**. At each time step t, the response of the synapse (**system**) to an artificial stimulation is recorded. This observation e_t is used by the **filter** to compute the posterior distribution of parameters $p(\theta|e_{1:t}, \Delta_{1:t})$. The **controller** then computes the next stimulation time Δ^*_{t+1} to maximize the expected gain of information. In classical experiment design, the stimulation times $\Delta_{1:T}$ are defined and fixed prior to the recordings.

be critical to the output of an experiment. However, most physiology experiments still rely on arbitrary and non-adaptive designs, which may not yield sufficient information about the studied system. Consequently, experiments often require more observations or repetitions to reach a certain result, which increases their cost, time, and need for subjects.

A relevant application for synaptic characterization is to determine when to optimally stimulate the presynaptic cell in order to maximize the informativeness of the recordings. As explained earlier, the parameters of a chemical synapse (namely, the number of presynaptic vesicles N, their release probability p, the quantal amplitude q, the short-term depression time constant τ_D, etc.) cannot be measured directly, but can be estimated from the synapse's postsynaptic responses to evoked stimuli. However, these estimates critically depend on the stimulation protocol being used. For instance, if inter-spike intervals are too large (i.e. longer than the plasticity time constants τ_D and τ_F), no short-term plasticity will appear in the recordings; conversely, a too high stimulation frequency will lead to a depletion of the presynaptic vesicles and to a poor informativeness of the postsynaptic currents.

An efficient theoretical framework to alleviate this issue is called Optimal Experiment Design (OED). OED selects experimental parameters to optimize a given output (for instance, to reduce the uncertainty of inferred parameters, or to increase the information gained per observation). It can thus be used to determine how to stimulate the presynaptic cell (i.e. to compute the ISIs $\Delta_{1:T}$). We will distinguish two kinds of Optimal Experiment Designs:

- **Optimal experiment design for parameters inference** (Section 1.8.1): for a given model $\mathcal{M}$, its goal is to optimize the accuracy of the estimates of the parameters θ of $\mathcal{M}$, i.e. to minimize the entropy of the posterior distribution $p(\theta|\mathcal{D}, \mathcal{M})$.
- **Optimal experiment design for model selection** (Section 1.8.2): its goal is to maximize the discriminability between competing candidate models, i.e. to minimize the entropy of $p(\mathcal{M}|\mathcal{D})$.

1.8.1 Optimal experiment design for parameter inference

In this setting, the objective of experiment design optimization is to minimize the variance of the estimates while reducing the cost of experimentation. Within OED, an interesting framework is Bayesian Active Learning (BAL). For a fixed model $\mathcal{M}$, the goal is to optimize the accuracy of the estimates of the parameters θ of $\mathcal{M}$, i.e. to minimize the entropy of the posterior distribution $p(\theta|\mathcal{D}, \mathcal{M})$. Since the model is fixed, we will drop the dependency on $\mathcal{M}$ in the notations. The utility $\mathcal{U}(\mathcal{D}, \Psi)$ of a given experimental protocol Ψ and of a data set $\mathcal{D}$ can be either defined as the gain in Shannon information between the prior and the posterior distribution of the parameters θ, as suggested in [78]:

$$\mathcal{U}(\mathcal{D}, \Psi) = \int d\theta \log p(\theta|\mathcal{D}, \Psi) p(\theta|\mathcal{D}, \Psi) - \int d\theta \log p(\theta) p(\theta) \tag{1.6}$$

$\mathcal{U}(\mathcal{D}, \Psi)$ can also be defined as the Kullback-Leibler divergence between the prior and the posterior:

$$\mathcal{U}(\mathcal{D}, \Psi) = D_{KL}(p(\theta|\mathcal{D}, \Psi)||p(\theta)) \tag{1.7}$$

The expected utility $\mathcal{U}(\Psi)$ of a protocol Ψ is finally the expected value of $\mathcal{U}(\mathcal{D}, \Psi)$ under $p(\mathcal{D}|\Psi)$, which yields the same result under (1.6) and (1.7):

$$\mathcal{U}(\Psi) = \int \int d\theta d\mathcal{D} \log p(\theta|\mathcal{D}, \Psi) p(\theta, \mathcal{D}|\Psi) - \int d\theta \log p(\theta) p(\theta) \tag{1.8}$$

which, by noting that $p(\theta|\mathcal{D}, \Psi) = \frac{p(\mathcal{D}|\theta, \Psi) p(\theta)}{p(\mathcal{D}|\Psi)}$, is equivalent to

$$\mathcal{U}(\Psi) = \int \int d\theta d\mathcal{D} \log p(\mathcal{D}|\theta, \Psi) p(\theta, \mathcal{D}|\Psi) - \int d\mathcal{D} \log p(\mathcal{D}|\Psi) p(\mathcal{D}|\Psi) \tag{1.9}$$

It is worth noting that $\mathcal{U}(\Psi)$ is actually the mutual information between $\mathcal{D}$ and θ. Indeed, (1.8) is equivalent to

$$\mathcal{U}(\Psi) = \int_\theta \int_\mathcal{D} \log p(\theta|\mathcal{D}, \Psi) p(\theta, \mathcal{D}|\Psi) + H(\theta) = H(\mathcal{D}|\Psi) + H(\theta) - H(\mathcal{D}, \theta|\Psi)$$

which yields $\mathcal{U}(\Psi) = I(\mathcal{D}; \theta|\Psi)$, i.e. the mutual information between $\mathcal{D}$ and θ for a given Ψ. Different Markov Chain Monte Carlo (MCMC) methods to compute $\mathcal{U}(\Psi)$ are described in [79]. Combining (1.8) and (1.9) yields

$$-\mathbb{E}_\mathcal{D}(H(\theta|\mathcal{D}, \Psi)) + H(\theta) = \int \int d\theta d\mathcal{D} \log p(\mathcal{D}|\theta, \Psi) p(\theta, \mathcal{D}|\Psi) + H(\mathcal{D}|\Psi)$$

which finally gives

$$H(\theta, \mathcal{D}|\Psi) = H(\mathcal{D}|\Psi) + \mathbb{E}_\mathcal{D}(H(\theta|\mathcal{D}, \Psi)) \tag{1.10}$$

The goal of optimal experiment design is to minimize the overall expected risk $\mathbb{E}_\mathcal{D}(H(\theta|\mathcal{D}, \Psi))$ (see Equation 1.8). If the left-hand side of 1.10 does not depend on Ψ (e.g. in normal linear regression), this is equivalent to maximizing the entropy of the

data $H(\mathcal{D}|\Psi)$, which is the Maximum Entropy Sampling (MES) principle described in [80].

In synaptic characterization, $\mathcal{D}$ corresponds to a set of t recordings $e_{1:t}$, Ψ corresponds to the associated stimulation times $\Delta_{1:t}$, and θ are the parameters of the chosen model $\mathcal{M}$. After having recorded t observations, BAL can be used to compute the optimal next stimulation time Δ^*_{t+1}, as illustrated in Figure 1.11. Once a new observation e_t has been obtained, the posterior distribution of parameters $p(\theta|e_{1:t}, \Delta_{1:t})$ can be computed. Based on this current estimation of the distribution of θ, Equation 1.8 can be used to compute Δ^*_{t+1}.

However, the applicability of this optimization scheme to real experiments is limited by two main drawbacks. Firstly, it requires performing high-dimensional integrations and optimizations in real time. Different methods have been proposed to compute Eq. 1.8: Monte Carlo (MC) methods [79] or a variational approach [81] can be employed, but they usually require long computation times that can be impractical if the time between successive experiments is short. Closed-form solutions only exist for some special cases, such as linear models or GLM [29]. Current methods are thus either too time consuming, or only applicable to specific models. Secondly, it only optimizes for the next stimulus input (an approach referred to as a myopic design), disregarding all future observations in the experiment. To address these issues, we propose in Chapter 5 an Efficient Sampling-Based Bayesian Active Learning (ESB-BAL) framework, which is efficient enough to be used in real-time biological experiments and to go beyond myopic approaches.

1.8.2 Optimal experiment design for model selection

Assuming M competing models $\{\mathcal{M}_i\}_{1\leq i\leq M}$, there are 3 ways to define an optimal experiment design for optimal model selection:

- Computation of one design Ψ^* to optimize the discriminability of the M models. It can be designed using the Jensen-Shannon divergence between the densities of the competing models, as suggested in [82]:

$$\Psi^* = \arg\max_{\Psi} \sum_{i=1}^{M} p(\mathcal{M}_i) D_{KL}\left(p(\mathcal{D}|\mathcal{M}_i, \Psi)||\sum_{i=1}^{M} p(\mathcal{M}_i)p(\mathcal{D}|\mathcal{M}_i, \Psi)\right) \quad (1.11)$$

- Computation of M designs $(\Psi^*_i)_{1\leq i\leq M}$ optimizing the discriminability of model $\mathcal{M}_i$ against the $M-1$ competing models. It can be done using a one-sided form of the Kullback-Leibler divergence, as suggested in [83]:

$$\Psi^*_i = \arg\max_{\Psi} \int d\mathcal{D} p(\mathcal{D}|\mathcal{M}_i, \Psi) \log\left(\frac{(1-p(\mathcal{M}_i))p(\mathcal{D}|\mathcal{M}_i, \Psi)}{\sum_{j\neq i} p(\mathcal{M}_j)p(\mathcal{D}|\mathcal{M}_j, \Psi)}\right) \quad (1.12)$$

- Computation of $M(M-1)$ designs $(\Psi^*_{i,j})$ aiming at maximizing the discriminability of $\mathcal{M}_i$ against $\mathcal{M}_j$:

$$\Psi^*_{i,j} = \arg\max_{\Psi}(D_{KL}\left(p(\mathcal{D}|\mathcal{M}_i, \Psi)||p(\mathcal{D}|\mathcal{M}_j, \Psi)\right)) \quad (1.13)$$

  None of these schemes has ever been used for synaptic characterization: in Chapter 6, we draw a link between our proposed definition of practical identifiability and Optimal Experiment Design for model selection.

Chapter 2

Aims of the thesis

The first aim is to perform a systematic review of existing methods for model fitting. An important aspect of synaptic characterization is the inference of the parameters characterizing a synapse (namely the number of presynaptic vesicles N, their release probability p, the quantal component q, the time constants for depression and facilitation τ_D and τ_F, etc). This can be done by fitting a generative model of postsynaptic currents on experimentally recorded EPSCs. Different model-based approaches are discussed in Chapter 3.

The second aim is to see how different statistical models can be fitted on experimental data. In model selection, different generative models of different complexities and having different features can be proposed. However, the possibility to detect the right model among a set of competing candidates will depend on the ground truth parameters of the synapse and on the experimental conditions: there will be regimes in which some features cannot be detected. We propose a new statistical tool for assessing the identifiability domain of a model in Chapter 4. Besides, classical model selection criteria are derived assuming that observations are independent, impeding their applicability to models with correlated data. We discuss the possibility to apply these criteria to a broader range of models in Appendix B.

The third aim is to study how model fitting can be improved by optimizing the inputs to the system so as to maximize the accuracy of the inferred parameters. Bayesian Active Learning (BAL) is an efficient framework for learning the parameters of a model, in which input stimuli are selected to maximize the mutual information between the observations and the unknown parameters. However, its applicability to real experiments is limited by its computational complexity. In Chapter 5, we introduce a new framework for efficient BAL, and apply it to synaptic parameters inference. We also explain how our proposed definition of practical identifiability can be linked to Optimal Experiment Design for model selection in Chapter 6.

Finally, the fourth aim is to make sense of the observed organization of synaptic proteins from a theoretical point of view. Microscopy observations have shown that presynaptic release sites and postsynaptic receptors are organized in ring-like patterns, which are disrupted upon genetic mutations. In Chapter 7, we propose a normative approach to this protein organization, and suggest that it might optimize a certain biological cost function (e.g. the mean current or SNR after vesicle release).

Chapter 3

Model-based inference of synaptic transmission

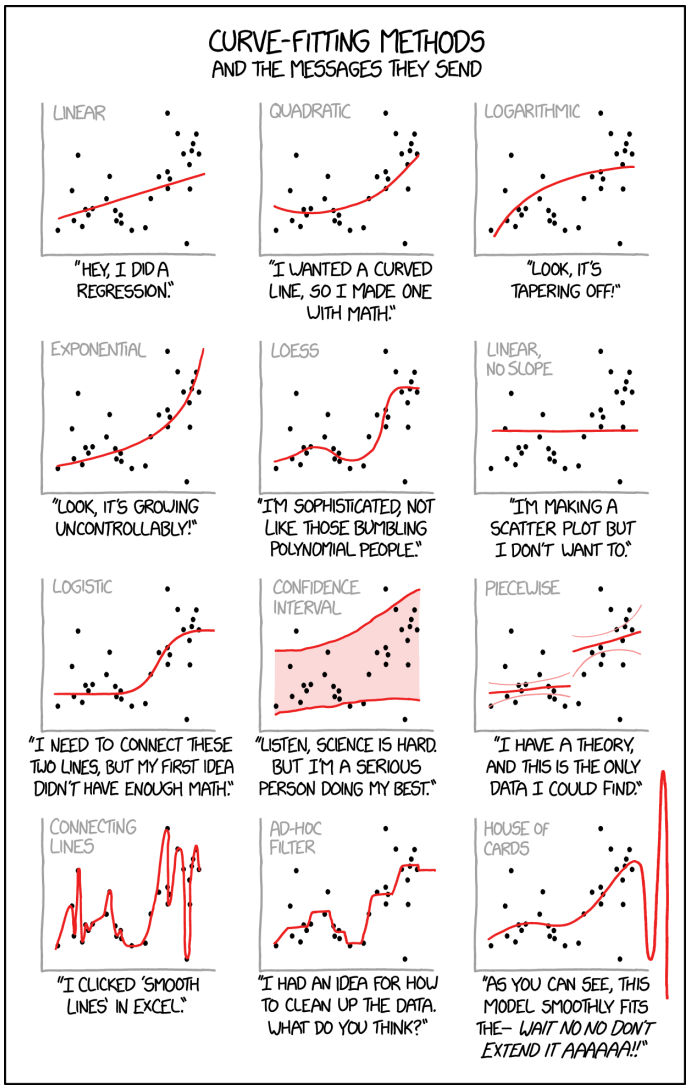


Figure 3.1: On thoughtful data fitting [1].

[1] `https://xkcd.com/2048/`, use with authorization.

This chapter includes the paper *Model-based inference of synaptic transmission* [3], published in Frontiers in Synaptic Neuroscience. The format was adapted to fit that of the thesis and the references have been included in the main bibliography.

Author contributions

Ola Bykowska, Camille Gontier, and Rui Ponte Costa generated the figures. Ola Bykowska, Camille Gontier, Anne-Lene Sax, David Jia, Milton Llera Montero, Alex Bird, Conor Houghton, Jean-Pascal Pfister, and Rui Ponte Costa wrote the manuscript.

3.1 Abstract

Synaptic computation is believed to underlie many forms of animal behavior. A correct identification of synaptic transmission properties is thus crucial for a better understanding of how the brain processes information, stores memories and learns. Recently, a number of new statistical methods for inferring synaptic transmission parameters have been introduced. Here we review and contrast these developments, with a focus on methods aimed at inferring both synaptic release statistics and synaptic dynamics. Furthermore, based on recent proposals we discuss how such methods can be applied to data across different levels of investigation: from intracellular paired experiments to in vivo network-wide recordings. Overall, these developments open the window to reliably estimating synaptic parameters in behaving animals.

3.2 Introduction

Modifications of synaptic transmission properties are believed to underlie learning, memory and, more generally, neural dynamics [84, 85, 86, 87, 39]. It is therefore of great importance to accurately infer synaptic transmission properties. Two key features that define synaptic communication are: stochastic transmission [88] and (relatively fast) temporal dynamics [89, 90]. The former is reflected as trial to trial variability of synaptic transmission as the combined result of pre- and postsynaptic sources of noise, such as probabilistic vesicle release (presynaptic) or binding of quantal neurotransmitter packets to (postsynaptic) receptors [91, 92]. Whereas temporal dynamics is reflected in the temporal modulation of synaptic responses, which is mediated by the multiple time constants of the synaptic transmission machinery. Such dynamics give rise to the commonly observed phenomenon of short-term plasticity (STP) [93, 90]. In this review we summarize, discuss and contrast recent developments in inference methods that capture either of these two elements (i.e., stochastic release and STP), or both. In particular our review focus on relatively simple phenomenological and statistical models, which abstract out the underlying biophysics and do not capture some aspects of synaptic transmission.

We also highlight recent advances toward inferring synaptic properties in vivo. Studying synaptic transmission parameters under naturalistic conditions is not only likely to give more precise parameters estimates, but also insights into what synaptic transmission properties are relevant in behaving animals [94, 95].

3.3 Inference of Stochastic Transmission

Synaptic transmission is inherently stochastic (see Fig. 3.2 for a schematic). In the quantal view of synaptic transmission neurotransmitter-containing vesicles (quanta) are released into the synaptic cleft from N release sites with probability P_{rel} [96, 97, 98] (Fig. 3.2A). Once released, neurotransmitters bind to postsynaptic receptors triggering a postsynaptic response with mean quantal amplitude q. A *binomial model* is often used to describe these three aspects (i.e. number of release sites N, release probability P_{rel} and the mean quantal amplitude q). In this model the mean peak of postsynaptic responses is given by $\mu = qNP_{\text{rel}}$ and their variance $\sigma^2 = q^2NP_{\text{rel}}(1 - P_{\text{rel}})$ (Fig. 3.2B) [2]. Several methods based on the binomial release model have been

[2]The binomial release model makes a few assumptions, namely that each site releases vesicles independently and that P_{rel} is the same across different release sites.

proposed to infer synaptic transmission parameters. A simple method relies solely on using the mean and variance to get estimates of both q and P_{rel} by rearranging the terms as $\hat{q} = \frac{\sigma^2}{\mu} + \frac{\mu}{N}$ and $\hat{P}_{\text{rel}} = \frac{\mu}{N\hat{q}}$ given a number of release sites N [99, 100]. The variance-mean analysis (also known as multiple-probability fluctuation analysis) is a slightly more advanced technique that relies on recording postsynaptic responses under different release probability conditions, which are set experimentally by varying the concentration of extracellular calcium. The relationship between the variance and the mean (i.e. μ and σ^2 as above) under different release probabilities is then fitted to the parabolic function given by the binomial model (Fig. 3.4A). This method estimates all three parameters (N, P_{rel} and q; see [98] for a detailed review on the topic). Because this method depends on having an accurate estimation of mean and variance of the postsynaptic responses, it requires relatively long and stable electrophysiological recordings under different conditions.

The mean-variance method described above may suffer from identifiability issues. For example, in the presence of a high level of noise it may not be possible to reliably identify the response amplitudes for the multiple peaks. In this case a simple Gaussian description of the synaptic responses may be preferable (Fig. 3.3A). In addition, these methods also rely on point estimates which may lead to inaccurate conclusions due to correlations in the parameters (see Fig. 3.3A for an example of such a case). A more principled approach to the problem that explicitly represents the uncertainty in the parameters should offer a better understanding of how well a particular model explains a given dataset.

[101] introduced a more principled quantal analysis method – Bayesian Quantal Analysis (BQA). This method applies Bayesian statistics which allows model inference to combine prior knowledge $P(M)$ over a model M with the data likelihood $P(D|M)$ following Bayes' theorem as $P(M|D) \sim P(M)P(D|M)$. In BQA, the prior is used to integrate *a priori* knowledge about the synaptic release statistics (e.g. expected bounds), which simultaneously models the distributions of postsynaptic responses recorded under multiple release probabilities (independent of each other). This is in contrast with standard mean-variance analysis described above, which simply models the mean responses across different release probabilities. By incorporating prior information, this method improves the accuracy of parameter inference and, importantly, reduces the number of samples needed compared to the mean-variance analysis (from about 100 samples to about 60 samples). Therefore, this new method may be preferable in experimental conditions where long recordings are particularly challenging (see a more detailed comparison in Table 1).

3.4 Inference of Short-Term Plasticity

Postsynaptic responses are dynamic – the peak response amplitude depends not only on the quantal parameters, but also on previous activity. If the presynaptic neuron fires in quick succession, the released vesicles are not given enough time to be recycled, which leads to less vesicles available for release. As a consequence synaptic responses become weaker, also known as *short-term depression* (Fig. 3.2A) and such recovery rates are often modelled with an exponential with τ_D. At the same time the presynaptic calcium levels can increase with every consecutive spike, which may lead to an increase in the postsynaptic response rather than a decrease – this is known as *short-term facilitation*.

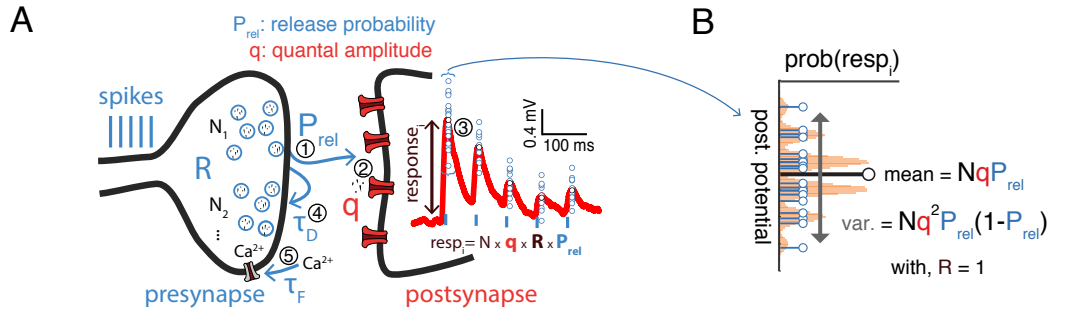


Figure 3.2: **Inference of synaptic transmission parameters.** (**A**) Schematic of synaptic transmission parameters. On the left the different elements of the synaptic transmission process are represented: first, presynaptic spikes (blue vertical bars) lead to release of vesicles containing neurotransmitter (R, for presynaptic resources) from one of N possible release sites with probability P_{rel}; second, released neurotransmitters (quanta) bind to postsynaptic receptors triggering a response with amplitude q ; third, this process triggers a postsynaptic response with average amplitude $NqRP_{\text{rel}}$, which takes into account both binomial and short-term synaptic plasticity; fourth, presynaptic vesicles are recovered with a time constant τ_D which may lead to *short-term depression* of consecutive postsynaptic responses (red trace on the postsynapse) before the presynaptic resources, R, fully recover; fifth, at the same time presynaptic voltage-dependent calcium (Ca^{2+}) channels can lead to calcium build-up on the presynapse (modelled by a time constant τ_F), which may increase release probability (P_{rel}) and in turn lead to an increase of consecutive postsynaptic responses, also known as *short-term facilitation* (not shown). (**B**) Postsynaptic responses exhibit variability (blue circles from (A) overlaid on top of the mean postsynaptic response in red). Such variability is often described as a simple binomial process , with N release sites and variance given by $Nq^2P_{\text{rel}}(1-P_{\text{rel}})$. Plot represents a binomial release model with $N = 5$, $P_{\text{rel}} = 0.5$ and some arbitrary q.

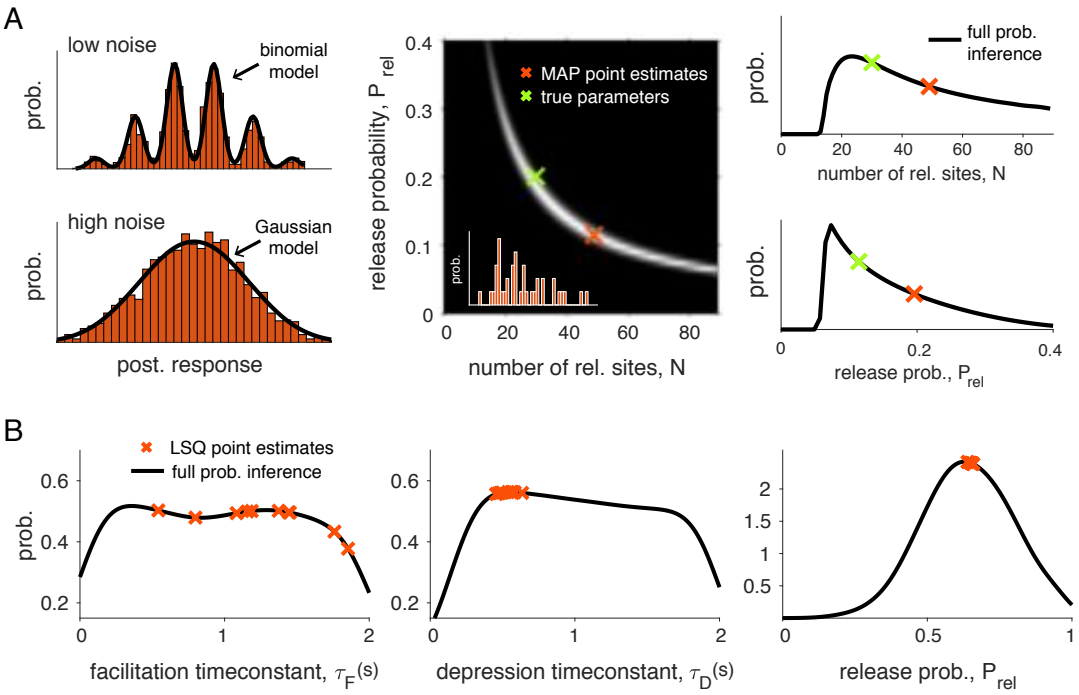


Figure 3.3: **Identifiability of synaptic transmission parameters.** (**A**) Identifiability issues of quantal release models. *Left upper figure*: Histogram of 2000 simulated postsynaptic responses with $N = 5$, $P_{\text{rel}} = 0.5$, $q = 1$, $\sigma = 0.3$. In this case it is possible to fit a binomial model. *Left lower figure*: same simulation, but for high noise ($\sigma = 0.7$). The quantal peaks (i.e. the parameter q) are not identifiable anymore if the recording noise is too high, and in this case a Gaussian model provides a better description of the synaptic responses. *Middle panel*: Pairwise posterior marginal for N and P_{rel} for a typical experimental case with 40 observations (simulated postsynaptic responses shown in inset) where the true parameters were $N = 30$, $P_{\text{rel}} = 0.2$ and $q = 1$ (green cross). The Maximum A Posteriori (MAP) estimates is obtained for $N = 49$ and $P_{\text{rel}} = 0.11$ (red cross): as N and p are anticorrelated, the posterior is roughly the same over a long band were N and p can be substituted, leading to inference error for a small number of observations. *Right panel*: Marginal posterior for N and P_{rel} from the previous panel. (**B**) Identifiability issues of short-term synaptic plasticity models. Given experimental data it is often of interest to infer the synaptic parameters. Two main types of inference have been applied: point estimations where a single scalar is estimated for one or more parameters (red crosses) or full probabilistic inference, where the full probability density over the parameters is obtained (black line). This particular example was obtained by inferring the Tsodyks-Markram model with four parameters given short-term plasticity recordings between pyramidal cells in layer-5 visual cortex (see [102] for more details, only three parameters are shown for simplicity: τ_F, τ_D and P_{rel}; cf. panel A). Point estimates were obtained using a standard least-square (LSQ) fitting method (simulated annealing). Full probabilistic inference was done using MCMC sampling following [102] (see main text for more details). As demonstrated by [102] the uncertainty over the parameters can be greatly reduced by using more informative protocols that cover a wider frequency range.

Approach	Binomial	STP	Inference quality	Experimental ease	Algorithm complexity
Mean-variance analysis [1]	✓	×	**	* (PSR)	$\mathcal{O}(M)$
Bayesian quantal analysis [2]	✓	×	***	** (PSR)	$\mathcal{O}(MN)$
Least-square STP fitting [3]	×	✓	**	*** (PSR)	$\mathcal{O}(M)$
Bayesian Gaussian-STP [4]	×	✓	****	*** (PSR)	$\mathcal{O}(MS)$
Binomial-STP [5]	✓	✓	***	*** (PSR)	$\mathcal{O}(MN^4)$
Bayesian binomial-STP [6]	✓	✓	*****	*** (PSR)	$\mathcal{O}(MN^4)$
Spike-based GLM [7]	×	✓	*	**** (spikes)	$\mathcal{O}(M)$

Table 3.1: Comparison of different model-based approaches. Note that the approaches that consider parameter uncertainty can be readily extended to Bayesian. [1] see [98] and Fig. 3.4A; [2] see [101] and Fig. 3.4A; [3] see for example [89, 103, 104, 105, 106, 107, 108, 109] and Fig. 3.4B; [4] see [102] and Fig. 3.4C; [5] see [110, 111] and Fig. 3.4C; [6] see [112] and Fig. 3.4C; [7] see [113] and Fig. 3.4D. In the $\mathcal{O}$ algorithm complexity analysis M refers to the number of data points, N to the number of release sites and S to the number of samples needed. Point estimate methods that obtain some measures of uncertainty of the parameters rely on getting multiple point estimates, whereas this comes naturally in full probabilistic methods (this is here reflected in the inference quality). The list of methods presented here is grouped into quantal methods (first two rows) and into STP models (last 5 rows) and then sorted by their publication date (earlier first). PSR: Postsynaptic responses. We use star-based ranking system for both inference quality and experimental ease, where one star means worse/harder.

3.4.1 Deterministic Models of Short-Term Plasticity

A number of deterministic short-term plasticity models have been proposed that characterise the dynamic properties of synaptic transmission (for a review on STP models see [114]). These models capture STP data relatively well, and thus may enable us to uncover how STP may be regulated under different conditions.

The parameters of these models are commonly fit using least-squares optimisation to obtain a single set of parameters (point estimates) where the goal is to find the best (or at least a good) set of parameters that captures a given experimental dataset [89, 103, 104, 105, 106, 107, 108, 109] (Fig. 3.4B).

However, estimating parameters of STP models poses a challenge. Similar to the issues highlighted above for binomial models, in most STP models different parameter sets produce model outputs that follow the observed data equally well (Fig. 3.3B; [102]). The existence of these multiple plausible solutions opens problems when relying on point estimates to draw conclusions about the underlying biological mechanisms. Therefore, it is important to also consider the uncertainty of the parameter estimation. Unlike single point estimate approaches, full probabilistic inference naturally captures parameter uncertainty. Note that this can also be in principle obtained using a sensitivity analysis when using standard fitting methods [115, 116], but these methods are often not principled and may not provide a complete picture of the parameter landscape. One form of probabilistic inference is full Bayesian inference where, similar to the BQA approach, we aim to obtain the posterior distribution of STP parameters given experimentally observed data.

[102] introduced the first Bayesian inference framework of STP models (Bayesian Gaussian-STP; Fig. 3.4C; Table 1). In this work the authors modelled the mean

postsynaptic peak responses using the Tsodyks-Markram STP model to account for the dynamic properties of the synapse [93, 89]. The Tsodyks-Markram STP model is a commonly used model built around the synaptic dynamics discussed above. In order to capture the variability of synaptic responses, [102] used a Gaussian approximation. The posterior distribution $P(M|D)$ was obtained by sampling of the parameter space using a Monte Carlo Markov Chain (MCMC) algorithm. MCMC methods rely on constructing a Markov chain [3] that should converge to the desired probability distribution in the equilibrium (i.e. after long enough observations).

This method was used to study the parameter uncertainty given datasets obtained with common experimental protocols. The posterior distributions revealed that some of the parameters from the Tsodyks-Markram STP models were poorly constrained by such experimental protocols (Fig. 3.3B). This observation led to the proposal of new experimental protocols that span a broader stimulation frequency range and result in substantially reduced uncertainty over the parameter values.

Furthermore, obtaining the posterior distribution helps to understand the dependencies between parameters, which is not straightforward using traditional fitting methods. For example, in [117], the authors used an MCMC method to obtain the posterior distribution over the parameters (similar to [102]) allowing the authors to highlight two strongly correlated parameters. Importantly, the identification of this correlation led to a reparameterization of the model which improved parameter inference. Therefore, obtaining the posterior distribution over the parameters makes it possible to characterise their uncertainty and explore possible dependencies between parameters. Such MCMC methods are relatively efficient as long as the model can be computed efficiently (up to a few seconds) and the number of STP parameters remains relatively low (less than a few dozens).

3.4.2 Stochastic Short-Term Plasticity Models

There are two important limitations of relying on deterministic STP models. First, the optimisation depends on an accurate estimation of the mean synaptic responses. As mentioned above, this requires a high number of trials, which is experimentally challenging (see Table 1). Second, by only considering averages these methods ignore the correlations between postsynaptic peaks, yet these correlations may provide valuable information to accurately infer the synaptic properties.

A couple of recent studies introduced methods that incorporate correlations between postsynaptic responses in the inference of STP parameters. These methods allow the extraction of both quantal and dynamic parameters of synaptic transmission from trains of postsynaptic responses without the requirement of averaging over multiple sweeps [110, 111, 112]. These studies implemented stochastic models (or generative models) of synaptic transmission by combining phenomenological Tsodyks-Markram STP models with binomial models of vesicular release and replenishment. The probability of vesicle release is derived from a Tsodyks-Markram model and the vesicle replenishment probability is modelled with a Poisson process controlled by a depression time constant parameter τ_D. The quantal size of the postsynaptic response evoked by each vesicle can be approximated by either a gamma distribution in [112], or an inverse Gaussian distribution in [111], and the total amplitude is modelled as a linear combination of these distributions. The choice of these distributions is mo-

[3] A Markov chain represents a probabilistic transition between states, in which a given transition depends only on the previous state.

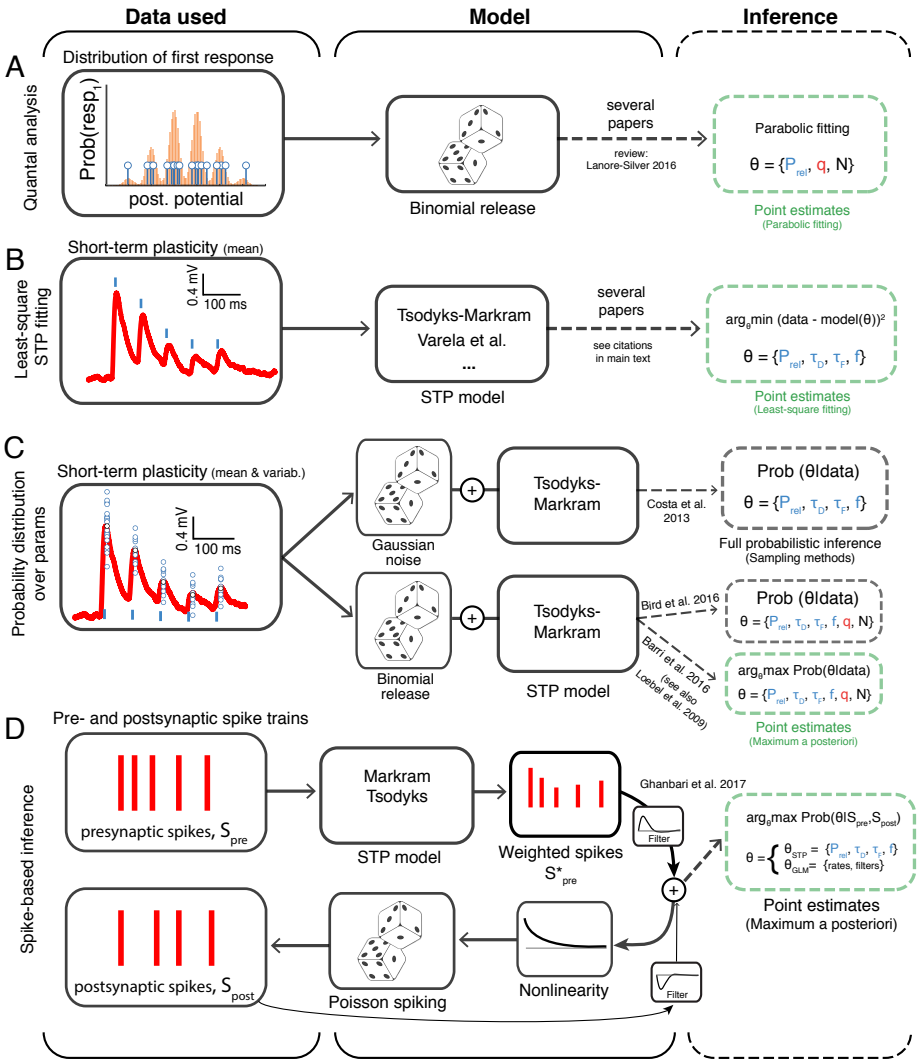


Figure 3.4: **Different approaches to model-based inference of synaptic transmission.** The different methods are organised based on the type of experimental data to which they are applied (first column), the model being assumed (second column) and the method of parameter inference (third column; dashed green and grey boxes indicate point estimate and full probabilistic inference, respectively). (**A**) Methods that use the variability of the first postsynaptic responses to infer binomial release statistics. (**B**) Methods that rely on multiple averaged responses to fit short-term plasticity (STP) models, which typically discard binomial release statistics. (**C**) Methods that directly consider both variability and multiple synaptic responses using probability theory to infer the synaptic transmission parameters. Here two variants have been explored: (i) a Bayesian framework where Gaussian noise is used to model the synaptic response variability [102] and (ii) a framework in which binomial release statistics are explicitly considered [112, 111]. The later has been explored using two variants: full inference (using sampling, [112]) and optimisation methods [111]. (**D**) Methods that work directly at the level of spike trains and try to infer short-term plasticity parameters. [113] introduced a new method based on generalised linear models (GLMs) to obtain point estimates of short-term plasticity models. P_{rel}: Release probability; q: mean quantal amplitude; N: number of release sites; τ_D: depression time constant; τ_F: facilitation time constant; f: facilitation rate. Similarly to previous figures the mean postsynaptic responses are shown in red, spikes in blue (vertical lines) and small blue circles represent individual samples of postsynaptic responses.

tivated by the fact that the quantal amplitude distribution is positively skewed, a feature that can not be captured by a Gaussian distribution.

The full formulation of the stochastic STP model allows to define the likelihood of the observed data given the model, $P(D|M)$. The stochasticity of the model introduces correlations between peaks in the train and these correlations pose the main difficulty in the likelihood calculation. In particular, because the amplitude of the postsynaptic response is dependent on the number of released vesicles. As discussed in [111], if the likelihood is to be formulated using the probability distribution of released vesicles, the number of terms in the calculation would grow exponentially. This becomes a permutation with repetition problem, in order to account for correlations of released vesicles the number of terms in the calculation would grow as $(N+1)^K$ with N being the number of release sites and K corresponding to the number of spikes in the train. To make the calculation more efficient, in both studies the likelihood function is formulated in terms of the probability distributions of the release sites before and after a spike (rather than continuously), which fully capture the state of the system.

These two studies apply different strategies to obtain point estimates from the likelihood. [111] uses an expectation-maximisation algorithm (referred to Binomial-STP in Table 1; see also [110]), while [112] uses MCMC sampling (referred to as Bayesian binomial-STP in Table 1). While both methods return a point estimate of the parameter set that maximises the likelihood function, only the sampling approach approximates the joint posterior distributions of the parameters. As discussed above, by obtaining the full likelihood, not just a point estimate, [112] explicitly quantifies the uncertainty over the parameters, and the full likelihood density can be analysed. Moreover, it also allows for correlations between the distributions over the parameters to be studied.

The main features of these approaches are (i) accounting for correlations between subsequent postsynaptic responses and (ii) using individual postsynaptic traces for fitting the models, which offers theoretical and practical advantages. Interestingly, both [111] and [112] report that considering correlations during inference yields estimates of synaptic parameters that are more accurate and require less sweeps when compared to ignoring correlations. This means that the experimental protocols can be shorter, hence making these inference methods particularly attractive for experiments *in vivo*.

3.5 Toward Inference of Synaptic Transmission in vivo

Recent developments have started to raise the possibility of accurately inferring synaptic transmission properties *in vivo*. One way to tackle this problem is to perform whole-cell recordings *in vivo* while stimulating the presynaptic neurons (or presenting a stimuli) [118, 100, 30]. This is a valuable approach that is enabling the community to confirm previous *in vitro* results *in vivo*. For example, [100] applied binomial-based estimation methods typically used in slices to *in vivo* data, and obtained results consistent with both modelling predictions and slice data. While [119] and [120] introduced new statistical methods with some success in inferring synaptic conductances from *in vivo* intracellular recordings and spike trains, respectively. However, these methods were not developed to estimate quantal or synaptic dynamics properties. In order to test how such synaptic features are shaped in more natural conditions across different brain regions new methods are required that can operate on the growing imaging-based or spike-based datasets.

One of the difficulties in inferring synaptic parameters from spikes is that several non-synaptic variables can have an impact on the spiking statistics. For example, when a presynaptic neuron fires at high frequencies one would expect a reduction in the firing rate of the postsynaptic neuron due to short-term depression at their synaptic connections, but a similar effect can also be mediated by postsynaptic neuron-wide adaptation mechanisms [121].

A first attempt at tackling this problem has recently been put forward [113]. In this framework, the authors extended a generalised linear model to infer both neuronal and STP parameters directly from spike-trains (Fig. 3.4D, referred to as spike-based GLM in Table 1). Interestingly, using their framework [113] showed that in a reduced system – a single postsynaptic neuron in slices with simulated inputs – postsynaptic adaptation can be distinguished from short-term depression as they are predominantly correlated with pre- and postsynaptic firing rates, respectively. More recently the same authors [122] went further and used their framework to show that functional connectivity with STP may explain the diversity of activity patterns observed *in vivo* between different brain areas. However, for these approaches to provide accurate estimates of synaptic transmission properties (Table 1) *in vivo* many other factors need to be considered in future work, such as network dynamics, cell-type specificity and dendritic integration.

3.6 Discussion

In this review we have provided an overview of standard methods and recent developments of model-based inference of synaptic transmission. From methods that rely on the binomial statistics of the first postsynaptic response alone to methods that consider the dynamics of consecutive synaptic responses (short-term plasticity) but also their statistical properties. Historically, inference methods have mostly focused on point estimations, which give a biased interpretation of synaptic data [102]. More recent developments have focused on full probabilistic inference, thus providing a more comprehensive picture on the most likely synaptic transmission parameters [101, 102, 112].

One research direction that should improve the inference quality of the short-term plasticity parameters is to optimise the experimental protocol, namely the timings of the presynaptic action potentials. The stimulation protocol needs to be within some acceptable range (a too high stimulation frequency would induce long-term plasticity and thereby violate the stationarity assumption). However, within such a range, there is a lot of freedom that can be exploited to improve the quality of the parameter estimates. For example, [102] explored a few different protocols (regular spike trains, regular spike train + recovery spike(s) or Poisson spike trains). It would be important to systematically study the space of protocols and determine which ones are the most informative. Pushing this idea even further, it would be interesting to design a closed-loop inference scheme such that after each spike and its subsequent postsynaptic response, the algorithm determines the best interval for the next spike that it is maximally informative about the synaptic parameters.

In this review, we have not covered some other properties that are of interest. One that has received attention recently is the inference of the size of the presynaptic readily-releasable vesicle pool [123, 124]. Additionally, we have focused on the binomial release model, but many synapses require different release probabilities and quantal amplitudes across release sites, which is better captured by multinomial statistics

[125, 98]. In future work, it would be important to understand how the developments reviewed here can also consider these finer aspects of synaptic transmission.

There have been remarkable developments in measuring synaptic properties with high temporal and spatial resolution [126, 127]. Of particular interest are recent advances in ultrafast optical glutamate sensors, which are enabling measurements of synaptic release with high accuracy [128]. These developments, when coupled with the statistical inference frameworks reviewed here ([102, 112, 113], but see also [129]), raise the possibility of accurate optical estimation of synaptic transmission properties in awake behaving animals.

Finally, there has been a recent surge in new and exciting large-scale recordings, such as voltage and calcium imaging [130], but also multi-electrode spike recordings [131]. With such methods at hand now is the right time to start asking questions that bridge systems neuroscience and synaptic transmission properties. For example, how do the quantal properties of synapses change over multiple brain areas as animals learn a particular task.

Taken together these novel inference and experimental methods open the possibility of testing different theories put forward for the role of synaptic transmission in learning and memory [132, 100, 85, 39], but also their impact in pathological states [133].

3.7 Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3.8 Acknowledgments

We would like to thank Mark van Rossum and Magnus Richardson for useful feedback on this review. OB and A-LS work was supported by the Wellcome Trust Doctoral Training Programme in Neural Dynamics, Grant no. 105207/Z/14/Z and 215303/Z/19/Z, respectively. MM was supported by the European Research Council, grant no. 741134 to Jeff Bowers.

Chapter 4

Identifiability of a binomial synapse

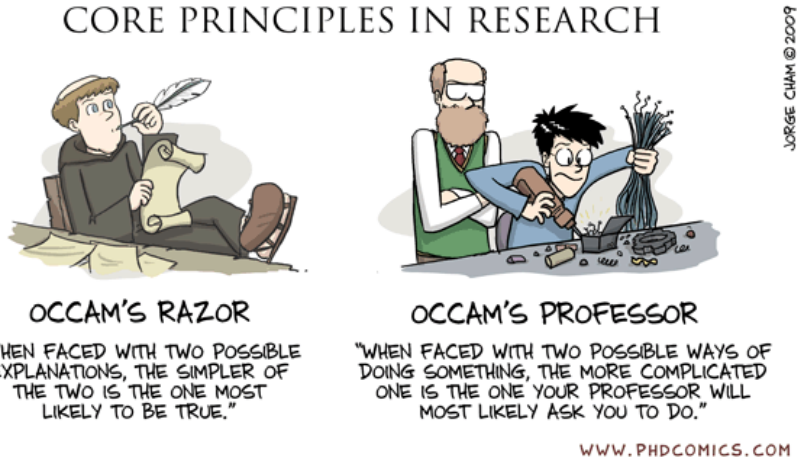


Figure 4.1: On correct model selection [1].

This chapter includes the paper *Identifiability of a binomial synapse* [134], published in Frontiers in Computational Neuroscience. Its abstract has been accepted for an oral presentation at the CNS conference 2019[2]. The format was adapted to fit that of the thesis and the references have been included in the main bibliography.

Author contributions

Camille Gontier and Jean-Pascal Pfister wrote the paper and derived the equations. Camille Gontier performed the numerical simulations, made the figures, and wrote the `MATLAB` code and the `Python` code. Jean-Pascal Pfister conceived of the presented idea and supervised the project. All authors contributed to the article and approved the submitted version.

[1] `https://phdcomics.com/comics/archive.php?comicid=1237`, use with authorization.

[2] Gontier C and Pfister J-P. Identifiability of a binomial synapse. CNS*2019, Barcelona, Spain. doi: `https://doi.org/10.1186/s12868-019-0538-0`

4.1 Abstract

Synapses are highly stochastic transmission units. A classical model describing this stochastic transmission is called the binomial model, and its underlying parameters can be estimated from postsynaptic responses to evoked stimuli. The accuracy of parameter estimates obtained via such a model-based approach depends on the identifiability of the model. A model is said to be structurally identifiable if its parameters can be uniquely inferred from the distribution of its outputs. However, this theoretical property does not necessarily imply practical identifiability. For instance, if the number of observations is low or if the recording noise is high, the model's parameters can only be loosely estimated. Structural identifiability, which is an intrinsic property of a model, has been widely characterized; but practical identifiability, which is a property of both the model and the experimental protocol, is usually only qualitatively assessed. Here, we propose a formal definition for the practical identifiability domain of a statistical model. For a given experimental protocol, this domain corresponds to the set of parameters for which the model is correctly identified as the ground truth compared to a simpler alternative model. Considering a model selection problem instead of a parameter inference problem allows to derive a non-arbitrary criterion for practical identifiability. We apply our definition to the study of neurotransmitter release at a chemical synapse. Our final contribution to the analysis of synaptic stochasticity is three-fold: firstly, we propose a quantitative criterion for the practical identifiability of a statistical model, and compute the identifiability domains of different variants of the binomial release model (uni or multi-quantal, with or without short-term plasticity); secondly, we extend the Bayesian Information Criterion (BIC), a classically used tool for model selection, to models with correlated data (which is the case for most models of chemical synapses); finally, we show that our approach allows to perform data free model selection, i.e. to verify if a model used to fit data was indeed identifiable even without access to the data, but having only access to the fitted parameters.

4.2 Introduction

Model selection is highly relevant to neuroscience, as neurons, dendrites, and synapses can be represented by models with different levels of complexity and abstraction. When it comes to fitting recorded data, predicting the output of a system to a given stimulus, or making sense of an observed phenomenon, several possible models can be used: this raises the question of what makes a good model. Finding the correct model is a crucial issue in studying the brain.

Firstly, a good model needs to be sufficiently complex to account for observed data, while being simple enough to generalize to future observations. Competing models are typically compared based on their ability to fit an observed data set, while being penalized for their complexity (or number of free parameters) to avoid overfitting. Different model selection tools (Bayesian Information Criterion, Akaike Information Criterion...) are classically used to determine which model is the best one to fit a data set [135].

Secondly, models also differ in their nature, and can be classified as phenomenological, normative, or biophysical. On the one hand, purely phenomenological models are useful for relating the output of a system to its input, and can provide a computationally efficient way to make prediction. However, as they are solely based on the

empirical relation between the input and the output of the system, and not on its inner biological principles, they lack interpretability. On the other hand, normative and biophysical models can be computationally challenging to fit on data, but are more realistic. In a normative approach, the output of a system is computed from an objective function which modelizes its high-level functions and principles. As opposed to this top-down approach, biophysical models aim at precisely describing the low-level features and inner biological principles of the system. An interesting property of these biophysical models is that their parameters correspond to real physical quantities: when the parameters of a system cannot be measured directly, they can be estimated by fitting a corresponding biophysical model on recorded output data of the system, a procedure known as model-based inference. By computing the likelihood of the data as a function of the parameters, it is possible to follow a maximum-likelihood approach to obtain a point estimate of the parameters [20], or to compute the full posterior distribution over them [5].

Such a parameter inference requires that the model used is identifiable. Structural (i.e. model-based) identifiability is a property of the model, regardless of experimental results. In a structurally identifiable system, the dimension of the output is sufficiently high with respect to the dimension of the parameters vector to uniquely define it: the parameters can be non-ambiguously inferred if the complete distribution of the output is known. Structural identifiability has been widely studied in many fields of physics and biology [70, 71, 72, 73, 74], and different criteria exist to assess the structural identifiability of a model [75].

This theoretical property is not equivalent to practical (i.e. experiment-based) identifiability, which is a property of both the model and the experimental protocol: a model which is structurally identifiable might lead to a poor practical identifiability of parameters if data points are noisy or scarce. The accuracy of model-based methods for inferring the values of parameters depends on the experimental protocol used to record the data, as observations need to be sufficiently informative to allow a correct estimation of the parameters. Contrary to structural identifiability, a quantitative criterion is lacking for practical identifiability, which is usually only qualitatively assessed. Non-practical identifiability refers to regimes in which parameters can only be loosely estimated; but one would need to define what does "loose" mean. Such a definition could be intrinsic to the model: a model could be considered as practically identifiable given a certain experimental protocol if the expected variance of its parameters' estimate is below a threshold. But this threshold would need to be arbitrarily defined. Here, we propose an extrinsic yet non-arbitrary definition of practical identifiability, by transforming a model identifiability problem into a model selection problem.

A model is said to be practically identifiable when its parameters can be correctly inferred given a certain experimental protocol. But, as explained previously, different possible models can be fitted on a data set. Recorded data need to be sufficiently informative not only to give a correct estimate of the parameters of a model, but also to select the correct model (i.e. the model from which they have been generated). We argue that a model is practically identifiable if and only if it is also correctly identified as the model providing the best fit to the data. For a given experimental protocol, we define the practical identifiability domain of a statistical model as the set of parameters for which the model is correctly identified as the ground truth compared to a simpler alternative submodel.

Our proposed definition of practical identifiability can be applied to any setting where

submodels or a nested family of models can be defined. Here, we apply it to the particular problem of estimating the parameters of a chemical synapse. A classical biophysical model used to describe the stochastic release of neurotransmitter at chemical synapses is called the binomial model [12], for which different variants of increasing complexity (in term of the number of free parameters) can be considered.

Different model-based approaches have been proposed [3] for obtaining an accurate estimate of the parameters describing a synapse (namely, its number of independent release sites, their release probability upon the arrival of a presynaptic spike, the quantum of current elicited by one release event, etc.) These parameters cannot be measured directly, but can be inferred using excitatory postsynaptic currents (EPSCs[3]) recorded on the post-synaptic side and elicited by experimental stimulation of the presynaptic cell. By measuring their values before and after a stimulation protocol, it is possible to study the mechanisms and loci of synaptic plasticity [137, 35, 36] and homeostasis [32, 33]. On a more theoretical level, a correct inference of synaptic parameters is necessary to study the computational role of synaptic stochasticity [31, 138]. Finally, an accurate inference of synaptic parameters would allow to clarify the role of synaptic transmission in different diseases [37], such as mental retardation [139], schizophrenia [140], Parkinson's disease [141], autism [142], Alzheimer's disease [143], compulsive behavior [144], and addiction [145].

Our final contribution to the analysis of synaptic stochasticity is three-fold. Firstly, we propose a definition for the practical identifiability of a model of synaptic transmission, and compute the identifiability domains of different variants of the binomial release model. Besides, we observe that model selection criteria are classically derived by assuming that recorded data are not correlated, which does not hold for most models of chemical synapse. We extend the Bayesian Information Criterion (BIC), a classically used tool for model selection, to models with correlated data. Finally, a proper description of the model selection step is often missing in studies where a model based-approach is used to infer synaptic parameters. We show that our approach allows to perform *data free* model selection, i.e. to verify if a model used to fit data was indeed identifiable even if a proper model selection step had not been performed.

4.3 Methods

4.3.1 Binomial models of neurotransmitter release

The classical binomial model

The quantal nature of synaptic transmission was first unveiled in [9], in which the authors observed that the postsynaptic responses to presynaptic stimulations were all multiples of a small unit of current. They explained how the total response is built up of several of these units, or quanta, each of them arising from a single presynaptic release event. Upon the arrival of an action potential in the presynaptic terminal, vesicles are released with a given probability p. The binomial model [12] assumes that there are N independent release sites and that for each site the release probability p is identical. Therefore, the number of released vesicles after spike i, k_i, is distributed according to a binomial distribution. This model further assumes that each vesicle release gives rise to a quantal current q, such that the overall excitatory postsynaptic

[3] It is also possible to perform model-based inference of synaptic parameters based on post-synaptic spike trains instead of EPSCs, as in [26, 136]

current is given by $e_i = qk_i + \epsilon$, where ϵ models a measurement noise typically drawn from a normal distribution of variance σ^2. Under the binomial model described by its parameters N, p, q, and σ, the distribution of EPSCs is given by

$$p(e_i) = \sum_{k_i=0}^{N} p(e_i|k_i)p(k_i)$$

where k_i follows a binomial distribution of parameters N and p, and e_i conditioned on k_i follows a normal distribution of mean qk_i and variance σ^2. Postsynaptic responses are characterized by their mean Npq and their variance $q^2Np(1-p) + \sigma^2$. A first feature of synaptic transmission is thus its stochasticity. Due to different sources of noise, such as probabilistic vesicles release or recording noise, postsynaptic recordings exhibit trial-to-trial variability.

Full model of synaptic transmission

Although this simple binomial model accounts for synaptic stochasticity, it does not allow to model its dynamics: postsynaptic responses do not only depend on the parameters of the synapse, but also on its previous activity. On the one hand, successive presynaptic stimulations within a short time interval will lead to a depletion of the readily-releasable vesicle pool, and hence to reduced successive postsynaptic responses, a phenomenon known as short-term depression. This can be modeled by assuming that the number of available vesicles at time i is $n_i \leq N$ (while the simplified binomial model described above assumes that all vesicles are readily releasable, and hence $n_i = N$). On the other hand, successive stimulations will gradually increase the presynaptic calcium concentration, and hence the release probability, which is called short-term facilitation.

Short-term depression and facilitation can be modeled using the Tsodyks-Markram model [17, 137]. It consists in two ordinary differential equations, which model the proportion of available vesicles r_i and the release probability u_i at time i. r_i is reduced by an amount $u_i r_i$ after each presynaptic spike, and recovers back to 1 with a depression time constant τ_D between each spike. Similarly, u_i is increased by an amount $p(1-u_i)$, and decays back to p (its baseline value) with a facilitation time constant τ_F. Different values of the parameters p, τ_D, and τ_F allow to represent different synaptic dynamics (either depression, facilitation, or no plasticity at all).

However, such a deterministic approach to short-term plasticity only allows to model averages, and neglects correlations between successive postsynaptic responses. In recent studies [20, 5], models of synapses incorporating both short-term plasticity and binomial models of vesicles release and refill have been proposed. In these models, the release probability u_i evolves according to the equation of the Tsodyks-Markram model, while each vesicle refills with a probability $1 - \exp(-\frac{\Delta_{t_i}}{\tau_D})$, where Δ_{t_i} is the time interval between two successive presynaptic stimulations. This approach allows to represent both the stochasticity and the dynamics of neurotransmitter release, and to compute the likelihood of a set of recorded data $\mathcal{D}$ given the parameters θ and the presynaptic stimulation protocol Ψ.

We consider a model of chemical synapse which encompasses both short-term depression (STD) and facilitation (STF) [20, 5]. Its parameters are (Figure 4.2 (**A**)):

- N: the number of independent release sites [-]
- p: their initial release probability [-]
- σ: the recording noise. It encompasses both the noise coming from the experimental apparatus (thermal noise of the amplifier, electric line noise, etc.) and from the recordings per se (such as fluctuations in the membrane potential of the cell) [A]
- q: the quantum of current elicited by one release event [A][4]
- τ_D: the time constant of vesicles refilling, and hence of short-term depression [s]
- τ_F: the time constant of Ca^{2+} dynamics, and hence of short-term facilitation [s]

which defines a vector $\theta = (N, p, q, \sigma, \tau_D, \tau_F)$. A probability conditioned on a parametrization θ is written p_θ.

$\mathbf{n} = \{n_i\}_{1 \leq i \leq T}$, $\mathbf{k} = \{k_i\}_{1 \leq i \leq T}$, and $\mathcal{D} = \{e_i\}_{1 \leq i \leq T}$ represent respectively the number of available vesicles at the moment of spike i, the number of vesicles released after spike i, and the i-th recorded EPSC (Figure 4.2 (**B**)). The experimental protocol $\Psi = \{t_1, t_2, \ldots, t_T\}$ encompasses the times of presynaptic stimulation: the time of the i-th spike is written t_i and $\Delta t_i = t_i - t_{i-1}$. For simplicity, we will drop the dependency on Ψ from the notations of probabilities.

The probability of recording $\mathcal{D}$ is computed as the marginal of the joint distribution of the observations $\mathcal{D}$ and the hidden variables $\mathbf{n}$ and $\mathbf{k}$:

$$p_\theta(\mathcal{D}, \mathbf{n}, \mathbf{k}) = p_\theta(e_1|k_1)p_\theta(k_1|n_1)p_\theta(n_1)\prod_{i=2}^{T} p_\theta(e_i|k_i)p_\theta(k_i|n_i)p_\theta(n_i|n_{i-1}, k_{i-1}) \quad (4.1)$$

where:

$p_\theta(e_i|k_i)$ is the emission probability, i.e. the probability to record e_i knowing that k_i vesicles released neurotransmitter and assuming a normally distributed recording noise[5]:

$$p_\theta(e_i|k_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(e_i - qk_i)^2}{2\sigma^2}\right) \quad (4.2)$$

$p_\theta(k_i|n_i)$ is the binomial distribution and represents the probability that, given n_i available vesicles, k_i of them will indeed release neurotransmitter:

$$p_\theta(k_i|n_i) = \begin{pmatrix} n_i \\ k_i \end{pmatrix} u_i^{k_i}(1 - u_i)^{n_i - k_i} \quad (4.3)$$

where the release probability u_i evolves as

[4]The outputs of a model of chemical synapse can be either electric postsynaptic currents (EPSC) or potentials (EPSP). In the latter case, σ and q will be expressed in [V] instead of [A].

[5]Other distributions can also be used for the emission probability. [20] assumed an inverse Gaussian to account for the observed right-skewness of mEPSP [146, 147].

$$u_i = p + u_{i-1}(1-p)\exp\left(-\frac{\Delta t_i}{\tau_F}\right) \tag{4.4}$$

with $u_1 = p$;

$p_\theta(n_i|n_{i-1}, k_{i-1})$ represents the process of vesicles refilling. During the time interval Δt_i, each empty vesicle can refill with a probability I_i:

$$p_\theta(n_i|n_{i-1}, k_{i-1}) = \begin{pmatrix} N - n_{i-1} + k_{i-1} \\ n_i - n_{i-1} + k_{i-1} \end{pmatrix} I_i^{n_i - n_{i-1} + k_{i-1}} (1 - I_i)^{N - n_i} \tag{4.5}$$

with

$$I_i = 1 - \exp\left(-\frac{\Delta t_i}{\tau_D}\right) \tag{4.6}$$

It is usually assumed that, at the beginning of the experiment, all release sites are filled, and hence that $n_1 = N$ [20, 5].

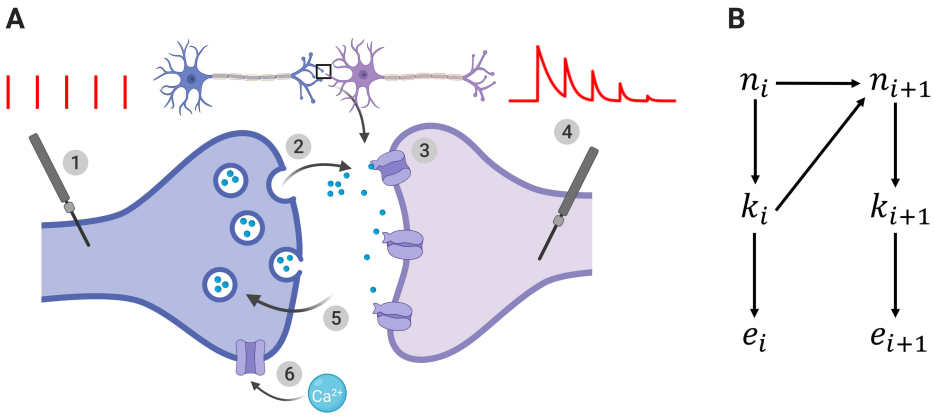


Figure 4.2: **(A)** Illustration of the binomial model. (1): the presynaptic button is artificially stimulated. Red vertical bars show 5 presynaptic spikes with a constant interspike interval. (2): these evoked stimuli lead to neurotransmitter release. After spike i, k_i vesicles (out of the n_i vesicles from the readily-releasable vesicle pool) release their neurotransmitter with a probability u_i. (3): neurotransmitter bind to receptors and elicit a postsynaptic current. A single release event triggers a quantal response of amplitude q. (4): the recorded postsynaptic response after spike i is the sum of the effects of the k_i release events. EPSCs correspond to the amplitude of each peak of the postsynaptic response to a presynaptic spike. (5): out of the N release sites, only n_i are in the readily-releasable vesicle pool at the moment of spike i. After releasing, vesicles recover with a time constant τ_D which determines short-term depression. (6): in the same time, each spike increases the calcium concentration in the presynaptic button, and hence increases the release probability u_i. This short-term facilitation is characterized by a time constant τ_F. **(B)** Generative model for the dynamical binomial model where n_i is the number of ready releasable vesicles and k_i is the number of released vesicles at time i. e_i is the EPSC amplitude at time t_i.

4.3.2 Models, submodels, and nested families

Definition 4.3.1. Model. For a given data set $\mathcal{D}$ and experimental protocol Ψ, a model $\mathcal{M}$ is defined as a triplet $\mathcal{M} = \{\Theta, \pi, \mathcal{L}\}$ where Θ is the space of parameters $\theta \in \Theta$, π is the prior for the parameters $\pi(\theta) = p(\theta|\mathcal{M})$, and $\mathcal{L}$ is the likelihood of the parameters $\mathcal{L}(\theta|\mathcal{D}) = p(\mathcal{D}|\theta, \mathcal{M}, \Psi)$.

Examples: Different models can be derived from equations (4.1) to (4.6). We consider 4 models of decreasing complexity:

Model $\mathcal{M}_3$ is the full model with both STD and STF. Its 6 parameters are N, p, q, σ, τ_D, and τ_F, and hence $\Theta_3 = \mathbb{N}^* \times [0,1] \times (\mathbb{R}_+)^4$. Its likelihood function $\mathcal{L}_3$ is obtained by marginalizing out the hidden variables $\mathbf{n}$ and $\mathbf{k}$:

$$\mathcal{L}_3(\theta|\mathcal{D}) = \sum_{\mathbf{n},\mathbf{k}} p_\theta(\mathcal{D}, \mathbf{n}, \mathbf{k}) \tag{4.7}$$

where $p_\theta(\mathcal{D}, \mathbf{n}, \mathbf{k})$ is given by Eq. (4.1).

Model $\mathcal{M}_2$ only has short-term depression (and no short-term facilitation). Its 5 parameters are N, p, q, σ, and τ_D, and hence $\Theta_2 = \mathbb{N}^* \times [0,1] \times (\mathbb{R}_+)^3$. Its likelihood $\mathcal{L}_2$ can be derived from (4.7) by further assuming that u_i is a constant equal to p.

Model $\mathcal{M}_1$ shows no short-term plasticity at all, and can be derived from model $\mathcal{M}_2$ by further assuming that I_i (defined in (4.6)) is a constant equal to 1 (and hence $n_i = N$). Its 4 parameters are N, p, q, and σ, and hence $\Theta_1 = \mathbb{N}^* \times [0,1] \times (\mathbb{R}_+)^2$. In this setting, data points are independent and (4.7) becomes

$$\mathcal{L}_1(\theta|\mathcal{D}) = \prod_{i=1}^{T} \left(\sum_{k_i=0}^{N} p_\theta(e_i|k_i) p_\theta(k_i) \right) \tag{4.8}$$

with $p_\theta(k_i) = \begin{pmatrix} N \\ k_i \end{pmatrix} p^{k_i} (1-p)^{N-k_i}$ being the binomial distribution;

Model $\mathcal{M}_0$ is a Gaussian model, in which EPSCs are simply generated from a normal distribution parametrized by its mean and variance. Its 2 parameters are μ and σ^2, and hence $\Theta_0 = \mathbb{R} \times \mathbb{R}_+$. Its likelihood $\mathcal{L}_0$ is simply

$$\mathcal{L}_0(\theta|\mathcal{D}) = \prod_{i=1}^{T} p_\theta(e_i) \tag{4.9}$$

with $p_\theta(e_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(e_i-\mu)^2}{2\sigma^2}\right)$.

To ensure the completeness of the definition of the models, we will assume for each parameter θ a uniform prior between two values $\theta_{\min}$ and $\theta_{\max}$ [5]. Note however that the approximate identifiability domain defined in (4.17) does not depend on the prior.

Definition 4.3.2. Submodels. Although ubiquitous in statistics (as in the likelihood-ratio test or Pearson's chi-squared test), the notion of submodels (or nested models) is rarely formally defined in the literature [148]. It is usually said that $\mathcal{M}_0$ is a submodel of $\mathcal{M}_1$ (or is nested within $\mathcal{M}_1$) if $\mathcal{M}_0$ can be obtained by constraining the parameters of $\mathcal{M}_1$ [149]. We propose the following formal definition, that encompasses the space of parameters, their priors, and their likelihood.

$\mathcal{M}_0 = \{\Theta_0, \pi_0, \mathcal{L}_0\}$ is said to be a submodel of $\mathcal{M}_1 = \{\Theta_1, \pi_1, \mathcal{L}_1\}$ if

1. $\Theta_0 \subset \Theta_1$ (i.e. the parameters of $\mathcal{M}_0$ also appear in $\mathcal{M}_1$)
2. $\pi_0(\theta_0) = \int_{\Theta_1 \setminus \Theta_0} \pi_1(\theta_0, \tilde{\theta}) d\tilde{\theta} \quad \forall \theta_0 \in \Theta_0$ (i.e. $\mathcal{M}_0$ and $\mathcal{M}_1$ share the same priors for the parameters they have in common)
3. $\forall \theta_0 \in \Theta_0 \ , \exists \tilde{\theta}$ s.t. $p(\mathcal{D}|\theta_0, \mathcal{M}_0) = p(\mathcal{D}|(\theta_0, \tilde{\theta}), \mathcal{M}_1)$ with $(\theta_0, \tilde{\theta}) \in \Theta_1$ (i.e. $\mathcal{M}_0$ can be retrieved from $\mathcal{M}_1$ by constraining its parameters).

We use the notation $\mathcal{M}_0 \preceq \mathcal{M}_1$.

Examples: The model $\mathcal{M}_2$ with only short-term depression is a submodel of $\mathcal{M}_3$ (which accounts for both depression and facilitation). Indeed, they have the parameters N, p, q, σ, and τ_D in common, and $\mathcal{M}_2$ can be retrieved from $\mathcal{M}_3$ by constraining $\tau_F \to 0$. Similarly, the model without STP $\mathcal{M}_1$ is a submodel of $\mathcal{M}_2$ where $\tau_D \to 0$; and the uni-quantal model $\mathcal{M}_0$ is a submodel of the multi-quantal model $\mathcal{M}_1$ where $p = 1$ and $\mu = Nq$.

We propose the following definitions to characterize the nestedness of a family of models:

Definition 4.3.3. Nested family. $\mathcal{F} = \{\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_n\}$ is said to be a nested family if

$$\mathcal{M}_i \preceq \mathcal{M}_j, \ \forall 0 \leq i \leq j \leq n$$

4.3.3 Structural identifiability

Definition 4.3.4. Structural identifiability domain. Consistently with [70, 75], let the structural identifiability domain Θ_S of a model $\mathcal{M} = \{\Theta, \pi, \mathcal{L}\}$ be defined as:

$$\Theta_S = \{\theta \in \Theta \mid (\forall \theta' \in \Theta) \ \theta \neq \theta' \iff p(\mathcal{D}|\theta, \mathcal{M}) \neq p(\mathcal{D}|\theta', \mathcal{M})\} \tag{4.10}$$

Similarly, $\mathcal{M}$ is said to be structurally identifiable if $\Theta = \Theta_S$.

If θ is in the structural identifiability domain of $\mathcal{M}$, it can be uniquely identified from $p(\mathcal{D}|\theta, \mathcal{M})$. For instance, a Gaussian distribution of mean μ and variance σ^2 is uniquely defined by its parameters $\theta = (\mu, \sigma^2)$. Its structural identifiability domain is thus $\Theta_S = \mathbb{R} \times \mathbb{R}^+$. Similarly, if $N \neq 0$, $p \neq 0$, $p \neq 1$, and $q \neq 0$, the probability density of EPSCs under the binomial model without short-term plasticity $\mathcal{M}_1$ is a mixture of Gaussian and is, as such, a structurally identifiable model if we restrict Θ_1 to $\mathbb{N}^* \times]0, 1[\times (\mathbb{R}_+^*)^2$ [150] (Figure 4.3).

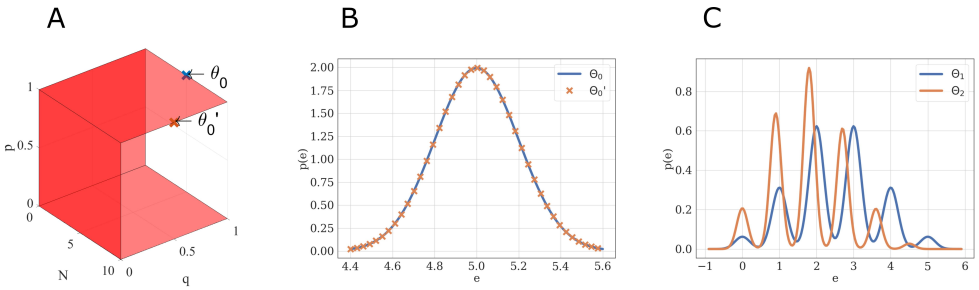


Figure 4.3: (**A**) Structural identifiability domain of the binomial model $\mathcal{M}_1$. This model, parametrized by $\theta = (N, p, q, \sigma)$, is structurally identifiable if and only if $N \neq 0$, $p \neq 0$, $p \neq 1$, and $q \neq 0$. These conditions are represented by the red hyperplanes in the $N \times p \times q$ domain. (**B**) Two different sets of parameters θ_0 and θ_0' may lead to the same distribution of observations if taken out of the structural identifiability domain. If $p = 1$, the distribution of EPSCs under model $\mathcal{M}_1$ follows a Gaussian law of variance σ^2 and mean Nq: different combinations of N and q can thus ambiguously describe it. Blue distribution: $N = 5$, $p = 1$, $q = 1$, $\sigma = 0.2$. Orange distribution: $N = 10$, $p = 1$, $q = 0.5$, $\sigma = 0.2$. (**C**) Two different sets of parameters θ_1 and θ_2 will lead to different distributions when taken within the structural identifiability domain of $\mathcal{M}_1$: its distribution is uniquely defined by its parameters. Blue distribution: $N = 5$, $p = 0.5$, $q = 1$, $\sigma = 0.2$. Orange distribution: $N = 5$, $p = 0.4$, $q = 0.9$, $\sigma = 0.15$.

4.3.4 Informative domain

In some regimes, parameters may not be precisely inferred from observations, even though the model is otherwise structurally identifiable. Indeed, in practice we usually only have access to a finite number of (possibly noisy) observations. *Practical* identifiability thus differs from the *structural* identifiability defined in Section 4.3.3.

A definition for the practical identifiability of a parameter has previously been proposed in [70], along with an approach for detecting practical non-identifiabilities based on the profile likelihood [76, 77]. The authors first define the likelihood-based confidence intervals for the estimator $\hat{\theta}_i$ of the i-th parameter of a model $\mathcal{M}$:

$$C_{i,\Delta} = \{\theta_i \mid \mathcal{L}(\hat{\theta}_i|\mathcal{D}) - \mathcal{L}(\theta_i|\mathcal{D}) < \Delta\}$$

where

$$\mathcal{L}(\theta_i|\mathcal{D}) = \max_{\theta_{j\neq i}} \mathcal{L}(\theta|\mathcal{D})$$

for a given threshold Δ. Then, they propose the following definition: *A parameter estimate $\hat{\theta}_i$ is practically non-identifiable, if the likelihood-based confidence region is infinitely extended in increasing and/or decreasing direction of θ_i, although the likelihood has a unique minimum for this parameter*, meaning that the decrease in likelihood compared to the optimal parameters estimate stays below the threshold Δ in direction of θ_i. When plotting the likelihood as a function of the parameters, a practical non-identifiability can be seen as an infinitely extended flat valley, in which the decrease in likelihood stays below Δ. The authors also describe an algorithm for computing the profile likelihood and hence detecting such practical non-identifiabilities:

Structural non-identifiable parameters are characterized by a flat profile likelihood. The profile likelihood of a practically non-identifiable parameter has a minimum, but is not excessing a threshold Δ for increasing and/or decreasing values of θ_i (see Figure 3 in [70]).

A first limitation of this definition is to be data-dependent: it only holds for a specific set of recorded data $\mathcal{D}$. Indeed, likelihood-based confidence intervals, and hence practical identifiability, are defined with respect to a certain data set $\mathcal{D}$, and may thus vary for different realizations of the experiment. However, an identifiability criterion can be made data-independent by averaging it over all possible realizations of $\mathcal{D}$, i.e. by computing its expectation with respect to the distribution $p(\mathcal{D}|\theta^*, \mathcal{M}, \Psi)$. Such an averaged criterion would correspond to the *a priori* expected identifiability before a specific $\mathcal{D}$ is recorded.

Practical information about θ is a function of the experimental protocol Ψ: for a given Ψ, the informative domain $\Theta_{\mathrm{I}}(\Psi)$ of a model $\mathcal{M}$ could be defined based on the variance of the estimator. For instance, in a Bayesian setting, the domain $\Theta_{\mathrm{I}}(\Psi)$ could be the set of parameters for which the expected informativeness of the posterior distribution of the parameters (measured as the Kullback-Leibler divergence between the posterior and the prior) is above a threshold Δ:

$$\Theta_{\mathrm{I}}(\Psi) = \{\theta^* \in \Theta \mid \langle D_{KL}(p(\theta|\mathcal{D}, \mathcal{M}, \Psi) \mid\mid p(\theta|\mathcal{M}))\rangle_{p(\mathcal{D}|\theta^*, \mathcal{M}, \Psi)} \geq \Delta\} \tag{4.11}$$

Although data-independent, this definition suffers from the same limitation as the one proposed in [70]: it requires to set a specific threshold Δ. Instead of defining an arbitrary criterion Δ on the possible precision of parameters estimate, we will derive our definition from a model selection argument.

4.3.5 Model selection

In model selection, the plausibility of two competing models $\mathcal{M} = \{\Theta, \pi, \mathcal{L}\}$ and $\mathcal{M}' = \{\Theta', \pi', \mathcal{L}'\}$ based on observations $\mathcal{D}$ can be assessed using the Bayes Factor [151]:

$$B_{\mathcal{M},\mathcal{M}'}(\mathcal{D}) = \frac{p(\mathcal{D}|\mathcal{M})}{p(\mathcal{D}|\mathcal{M}')} = \frac{\int_\Theta \mathcal{L}(\theta|\mathcal{D})\pi(\theta)d\theta}{\int_{\Theta'} \mathcal{L}'(\theta|\mathcal{D})\pi'(\theta)d\theta} \tag{4.12}$$

If the Bayes Factor is superior to 1, then the evidence for $\mathcal{M}$ is higher than the evidence for $\mathcal{M}'$. It is worth pointing out that the Bayes Factor will not only favor models which provide a good fit to the data, but also includes a tendency to favor simpler models, a natural form of Occam's Razor [152, 68]. Indeed, a complex model (i.e. a model with many independent parameters or with a broader prior for its parameters) will be able to explain a larger set of possible observed data than a simple model; but this comes at the price of spreading its likelihood over a larger set of possible outcomes. Hence, if two models fit the observed data equally well, the simpler one will be favored.

4.3.6 Proposed definition of practical model identifiability

To compute the identifiability domain of any model $\mathcal{M}$ compared to another model $\mathcal{M}'$, we introduce the Average Log Bayes Factor:

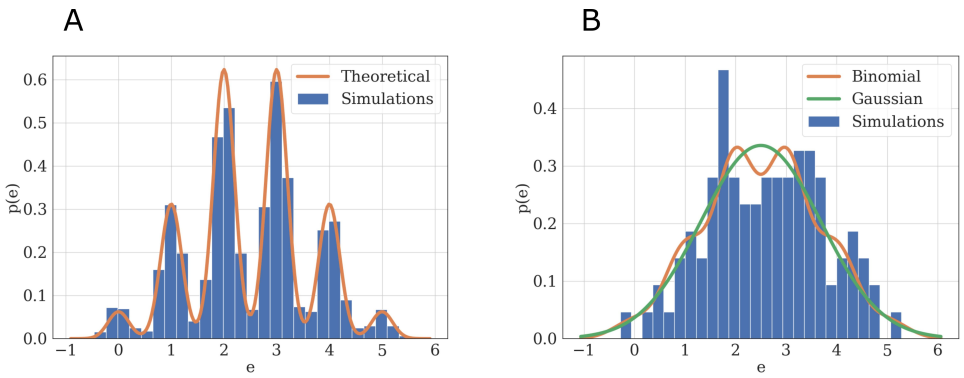


Figure 4.4: (**A**) Illustration of practical identifiability. Orange: theoretical distribution for model $\mathcal{M}_1$ parametrized with $N = 5$, $p = 0.5$, $q = 1$, $\sigma = 0.2$. Blue: histogram of 2000 simulated EPSCs generated using the same parameters. Parameters can be precisely inferred from the observations, which fit their theoretical distribution. (**B**) Illustration of practical non-identifiability. Blue: histogram of 100 simulated EPSCs with $N = 5$, $p = 0.5$, $q = 1$, $\sigma = 0.4$. Due to the small number of data points and high recording noise σ, the binomial parameters can only be loosely estimated, which is characterized by the fact that a Gaussian distribution (green) will provide a better fit to the data than a binomial distribution (orange).

$$\mathcal{B}_{\mathcal{M},\mathcal{M}'}(\theta^*, \Psi) = \langle \log B_{\mathcal{M},\mathcal{M}'}(\mathcal{D}) \rangle_{p(\mathcal{D}|\theta^*,\mathcal{M},\Psi)} \tag{4.13}$$

For a given parameter θ^* and protocol Ψ, model $\mathcal{M}$ is said to be practically identifiable compared to $\mathcal{M}'$ if $\mathcal{B}_{\mathcal{M},\mathcal{M}'}(\theta^*, \Psi) \geq 0$. Intuitively, the identifiability domain of $\mathcal{M}$ compared to $\mathcal{M}'$ corresponds to all the settings (parameters and protocols) for which, on average, data generated from the ground truth $\mathcal{M}$ will be better explained by $\mathcal{M}$ than by $\mathcal{M}'$.

In contrast to the definition in [70], our proposed definition does not require to set a (possibly arbitrary) threshold Δ. Instead, it is derived from a model selection criterion. We argue that a necessary and sufficient condition for the parameters of a model $\mathcal{M}$ to be practically identifiable is for $\mathcal{M}$ to be itself practically identifiable. In some settings (as for the nested models of chemical synapse described in Section 4.3.1), a family of submodels might naturally arise, while the choice of a threshold Δ would be arbitrary.

Another interest of our approach is to be data-independent, while the definition proposed in [70] only holds for a specific set of recorded data $\mathcal{D}$. Indeed, we define practical identifiability as a data-independent and intrinsic property of the model $\mathcal{M}$ and experimental protocol Ψ. As the log-Bayes Factor in (4.13) is averaged over all possible realizations of $\mathcal{D}$, it corresponds to the *a priori* expected identifiability before $\mathcal{D}$ is recorded. Our approach thus allows to define practical identifiability *domains*:

Definition 4.3.5. Practical identifiability domain. Consider a model $\mathcal{M} = \{\Theta, \pi, \mathcal{L}\}$ and a submodel $\mathcal{M}' = \{\Theta', \pi', \mathcal{L}'\}$ of $\mathcal{M}$. For a given experimental protocol Ψ, the practical identifiability domain $\Theta_{\mathrm{P}}(\Psi)$ of $\mathcal{M}$ is the set of parameters θ^* for which it is identifiable compared to its submodel:

$$\Theta_{\mathrm{P}}(\Psi) = \{\theta^* \in \Theta \mid \mathcal{B}_{\mathcal{M},\mathcal{M}'}(\theta^*, \Psi) \geq 0\} \tag{4.14}$$

Note that in the limit where the priors π and π' are highly peaked (i.e. $\pi(\theta) = \delta(\theta - \bar{\theta})$ and $\pi'(\theta) = \delta(\theta - \bar{\theta}')$), the condition $\mathcal{B}_{\mathcal{M},\mathcal{M}'}(\theta^*, \Psi) \geq 0$ is always satisfied due to Gibbs' inequality. In this case we have $\Theta_{\mathrm{P}}(\Psi) = \Theta, \forall \Psi$. However, generically the condition $\mathcal{B}_{\mathcal{M},\mathcal{M}'}(\theta^*, \Psi) \geq 0$ is not always satisfied since $p(\mathcal{D}|\mathcal{M})$ is not equal to $p(\mathcal{D}|\theta^*, \mathcal{M})$. The latter is the probability of observing $\mathcal{D}$ given $\mathcal{M}$ and a certain parametrization θ^*, while the former is the marginal likelihood over all parameters (4.12).

Two examples can illustrate this correspondence between model selection and parameter inference. Consider first the case of data recorded from $\mathcal{M}_1$. If the experimental protocol is not sufficiently informative (i.e. if data are scarce or noisy), not only will the inference of synaptic parameters be poor, but a Gaussian distribution will also provide a better fit than a binomial release model to the data. Indeed, as $[e_i|k_i] \sim \mathcal{N}(qk_i, \sigma^2)$, in the absence of recording noise (i.e. if $\sigma = 0$), the distribution of EPSCs is a serie of Dirac delta functions located at each multiple of the quantal size qk for $k \in \{0, 1, \ldots, N\}$. In this ideal case, q is clearly identifiable (Figure 4.4 (**A**)). However, upon addition of a recording noise σ, EPSCs are normally distributed around qk for $k \in \{0, 1, \ldots, N\}$, and the peaks on the histogram corresponding to each multiple of the quantal size might overlap if σ is sufficiently high with respect to q (Figure 4.4 (**B**)).

Similarly, we can consider the example of a synapse which shows short-term depression (STD) with a time constant τ_D (model $\mathcal{M}_2$). If the presynaptic cell is stimulated with an inter-spike intervals longer than τ_D, no depression will be visible in the recorded data, and the true model with STD will not be identifiable from a simpler binomial model without STD. In the same time, it will impossible to correctly infer the value of τ_D.

Our proposed definition of practical identifiability and of the identifiability domain of a model extend the landscaping technique introduced in [153] as well as the framework for testing identifiability of Bayesian models introduced in [154]. Especially, comparing the expected supports $\langle \log p(\mathcal{D}|\mathcal{M}) \rangle_{p(\mathcal{D}|\theta^*,\mathcal{M},\Psi)}$ and $\langle \log p(\mathcal{D}|\mathcal{M}') \rangle_{p(\mathcal{D}|\theta^*,\mathcal{M},\Psi)}$ of $\mathcal{M}$ and $\mathcal{M}'$ (given that values are averaged over $\langle \cdot \rangle_{p(\mathcal{D}|\theta^*,\mathcal{M},\Psi)}$) allows us to define a quantitative criterion for identifiability.

The model evidence $p(\mathcal{D}|\mathcal{M})$ in (4.12) is often intractable in practice for complex models, as it requires to integrate marginals for each parameter. Different methods have been proposed to approximate it: MCMC computations [155], Savage-Dickey method [156], supermodels [157]. A practical and time-efficient approximation of the model evidence is given by the Bayesian Information Criterion $\mathrm{BIC}_{\mathcal{M}}(\mathcal{D})$ [55]:

$$\mathrm{BIC}_{\mathcal{M}}(\mathcal{D}) \stackrel{!}{=} -2 \log \mathrm{p}(\mathcal{D}|\widehat{\theta}, \mathcal{M}) + \mathrm{k}_{\mathcal{M}} \log(\mathrm{T}) \approx -2 \log \mathrm{p}(\mathcal{D}|\mathcal{M}) \tag{4.15}$$

where $\hat{\theta} = \arg\max_\theta \mathcal{L}(\theta|\mathcal{D})$ is the maximum likelihood estimator (MLE) of $\mathcal{L}(\theta|\mathcal{D})$, $k_{\mathcal{M}} = \dim()$ is the number of independent parameters of $\mathcal{M}$, and $T = |\Psi|$ is the number of data points in $\mathcal{D}$. A detailed derivation is provided in Supplementary Material. The BIC is the sum of two terms: a likelihood term $-2 \log p(\mathcal{D}|\hat{\theta}, \mathcal{M})$ which represents the ability of the model $\mathcal{M}$ to explain $\mathcal{D}$, and a penalty term $k_{\mathcal{M}} \log(T)$ which favors simpler models, as explained in Section 4.3.5.

The BIC is commonly used as an approximation of the model evidence $p(\mathcal{D}|\mathcal{M})$ in model selection: the model with the lowest BIC is preferred over the others. The main advantage of using the BIC is to transform a complex integration problem (i.e. the computation of $p(\mathcal{D}|\mathcal{M})$) into a simpler optimization problem (i.e. the computation

of $\hat{\theta}$). Besides, it allows to perform model selection without the need to specify a prior for the parameters, and is thus a popular tool for model selection [135].

As stated in Supplementary Material, the approximation $\mathrm{BIC}_{\mathcal{M}}(\mathcal{D}) \approx -2\log \mathrm{p}(\mathcal{D}|\mathcal{M})$ is only valid under the hypothesis that data points are independent and identically distributed (i.i.d.), which is not the case for models with short-term plasticity. If data are correlated, we are left with the following approximation, which does not simplify in the general case:

$$-2\log p(\mathcal{D}|\mathcal{M}) \approx -2\log p(\mathcal{D}|\hat{\theta},\mathcal{M}) + \log(|H(\hat{\theta})|) \tag{4.16}$$

where $H(\hat{\theta})$ is the Hessian matrix of $-\log p(\mathcal{D}|\theta,\mathcal{M})$ in Eq. (4.15).

We emphasize that the classical definition of the BIC (4.15) should not be used if observations are correlated. Here, for models in which output are not independent, we use the approximation given by Eq. (4.16), in which the term $k_{\mathcal{M}}\log(T)$ in the BIC is replaced by $\log(|H(\hat{\theta})|)$. In some settings, the computation of the Hessian matrix can be challenging. However, MCMC methods can be used to approximate $H(\hat{\theta})$, even without an explicit expression for the gradient of the function [158]. In our case, a numerical method for computing $|H(\hat{\theta})|$ is detailed in the Supplementary Material.

Using approximation (4.15) in definition (4.14) yields the following approximation for the practical identifiability domain in case the model evidence $p(\mathcal{D}|\mathcal{M})$ in (4.12) is intractable:

$$\tilde{\Theta}_P(\Psi) = \{\theta^* \in \Theta \mid \langle \mathrm{BIC}_{\mathcal{M}}(\mathcal{D})\rangle_{\mathrm{p}(\mathcal{D}|\theta^*,\mathcal{M},)} \leq \langle \mathrm{BIC}_{\mathcal{M}'}(\mathcal{D})\rangle_{\mathrm{p}(\mathcal{D}|\theta^*,\mathcal{M},)}\} \tag{4.17}$$

4.4 Results

4.4.1 Identifiability domain of the binomial model without short-term plasticity

We study here the conditions under which a binomial model without short-term plasticity $\mathcal{M}_1$ can be correctly identified from a Gaussian model having the same mean and variance ($\mathcal{M}_0$). In order for the binomial model to be identifiable from a Gaussian quantum-less distribution, the recording noise needs to be sufficiently low compared to q for the peaks on the histogram of recorded EPSC to be identified. We will thus plot the identifiability domain as a function of the recording noise σ for a fixed q. The identifiability domain corresponds to the points θ in the parameters space Θ_1 for which the average BIC of $\mathcal{M}_1$ over all possible outputs of $\mathcal{M}_1$ parametrized with θ is lower than the average BIC of $\mathcal{M}_0$.

Per se, the identifiability domain depends on all the parameters of $\mathcal{M}_1$, as well as on the experimental protocol. For simplicity and in order to obtain a plot in 2 dimensions, we will only plot it as a function of p and σ while holding other variables to a fixed value. For a given experimental setup Ψ (which encompasses only the number of recorded data points T, the inter-spike intervals playing no role in these models), the following Markov-Chain-Monte-Carlo (MCMC) procedure is implemented:

1. A set of values p^* and σ^* are drawn from the space of possible values for p and σ;

2. Using p^* and σ^*, 400 independent data sets $(\mathcal{D}_i)_{1\leq i\leq 400}$ are generated from $\mathcal{M}_1$. Each data set consists in T EPSCs;

3. For each $\mathcal{D}_i$, the BIC of both models are computed; these values are averaged over i to compute an average BIC and identifiability is assessed if $\mathcal{M}_1$ is preferred over $\mathcal{M}_0$, which corresponds to the black dots in Figure 4.5 (**A**).

The procedure of plotting a complete identifiability domain can be quite time consuming. Indeed, it requires to span the entire space of parameters; for each vector θ^*, to generate a large number of independent data sets $(\mathcal{D}_i)$; and for each of these data sets, to compute the maximum likelihood estimator $\hat{\theta}$ using the Expectation-Maximization algorithm [20]. Details on the computation of $\hat{\theta}$ are available in Supplementary Material.

However, as both models $\mathcal{M}_0$ and $\mathcal{M}_1$ generate i.i.d. data, and by making the approximation $\hat{\theta} \approx \theta^*$ (i.e. by assuming that the maximum likelihood estimator $\hat{\theta}$ will be close to the true value θ^* from which data were generated), the condition that model $\mathcal{M}_1$ is identifiable (4.17) can be approximated as follows:

$$\begin{aligned} -2T \int p(e|\theta^*, \mathcal{M}_1) \log p(e|\theta^*, \mathcal{M}_1) de + k_{\mathcal{M}_1} \log(T) \leq \\ -2T \int p(e|\theta^*, \mathcal{M}_1) \log p(e|\hat{\theta}_{\mathcal{M}_0}, \mathcal{M}_0) de + k_{\mathcal{M}_0} \log(T) \end{aligned} \quad (4.18)$$

where $\hat{\theta}_{\mathcal{M}_0} = (\mu, \sigma^2)$ represents the mean $\mu \approx N^* p^* q^*$ and the variance $\sigma^2 \approx N^* p^* (1 - p^*) q^{*2} + \sigma^{*2}$ of the data generated from $\mathcal{M}_1$.

The condition specified by inequality (4.18) can be checked for any point θ^* without the need to generate a large number of independent data sets nor to compute the estimator $\hat{\theta}$. Solving (4.18) numerically for σ allows to draw the border of the identifiability domain of $\mathcal{M}_1$, represented as solide lines in Figure 4.5 (**A**) and (**B**).

Several points are worth highlighting. Firstly, Figure 4.5 (**A**) shows a good agreement between the results of the MCMC simulations (black dots) and those from the semi-analytical method (4.18) (blue line). Secondly, as expected, Figure 4.5 (**A**) and (**B**) illustrate that the identifiability domain increases with the number of data points T: intuitively, a larger data set facilitates the correct identification of a complex model. Besides, irrespective of the values of T and σ, for $p = 0$ and $p = 1$ the model $\mathcal{M}_1$ is structurally indistinguishable from a Gaussian distribution (see Figure 4.3 (**A**)). Finally, the maximum noise σ which makes the binomial model $\mathcal{M}_1$ indistinguishable from a Gaussian distribution $\mathcal{M}_0$ is larger for extreme values of p (close to 0.9 or 0.1) than for $p = 0.5$. Indeed, in the latter case, the distributions of EPSC will be symmetric (as in the upper panel of Figure 4.5 (**C**)), and hence just a little increase in recording noise will be enough to cover the inter-peak intervals and make the distribution Gaussian-shaped. In the former case, the distribution will be highly skewed, and thus difficult to approximate with a normal distribution.

The same approach can then be extended to more complicated models, by defining their identifiability domains as the part of the parameters plane where their average BIC will be lower than the BIC of a simpler one.

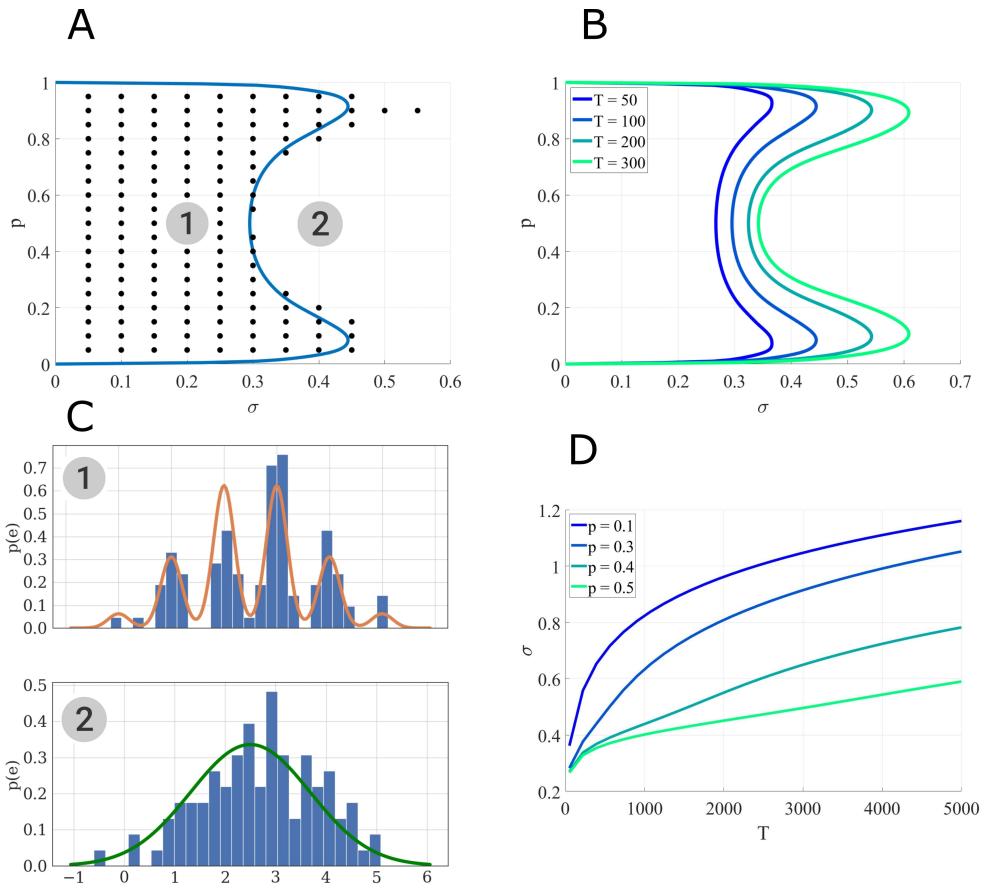


Figure 4.5: (**A**) Identifiability domain of $\mathcal{M}_1$ as a function of p and σ. Blue line: domain of identifiability from Eq. (4.18). On the left part of the blue line, the recording noise σ is sufficiently low to identify $\mathcal{M}_1$. Black dots: values (p^*, σ^*) for which the average BIC of 400 data sets results in the correct identifiability of $\mathcal{M}_1$. Results obtained for $N = 5$, $q = 1$, and $T = 100$. (**B**) Identifiability domain of the binomial model $\mathcal{M}_1$ compared to a Gaussian distribution $\mathcal{M}_0$, computed from (4.18), for different values of T. (**C**) To visualize the effect of σ on the data, this panel shows histograms of data generated from $\sigma = 0.2$ (1) and $\sigma = 0.4$ (2), alongside with their theoretical distribution from Eq. (4.8) (orange line in the upper panel) or when a Gaussian distribution is fitted on them (green line in the lower panel). In the identifiability domain (1), quantal peaks are clearly visible. Outside of the identifiability domain (2), the binomial distribution becomes Gaussian-shaped. (**D**) Another visualization of the identifiability domains displayed in (**A**) and (**B**). For different values of p, the maximum recording noise σ (i.e. the boundary of the identifiability domain) is plotted as a function of the number of data points T. The identifiability domain increases with T: intuitively, a larger data set facilitates the correct identification of a complex model.

4.4.2 Identifiability domain of the binomial model with short term depression

We study here the conditions under which a binomial model with short-term depression $\mathcal{M}_2$ can be correctly identified from a model without short-term plasticity ($\mathcal{M}_1$). In a first example, we assume that the presynaptic cell is stimulated at a constant inter-spike interval (ISI), which needs to be sufficiently short with respect to the time

constant τ_D to make depression visible. We thus plot the identifiability domain as a function of both p and τ_D. We use the same method as in 4.4.1: For each set of parameters p^* and τ_D^*, 400 independent data sets are generated from $\mathcal{M}_2$. Both models $\mathcal{M}_2$ and $\mathcal{M}_1$ are fitted on them, and black dots in Figure 4.6 (**A**) correspond to the parameters for which the average BIC of $\mathcal{M}_2$ is lower than the average BIC of $\mathcal{M}_1$.

As expected, we verify that the identifiability of $\mathcal{M}_2$ is only possible when τ_D is sufficiently long with respect to the inter-spike interval. Besides, if the release probability p is low, correlations between recordings will be weak and the effect of short-term depression will not be detectable. A major difference between models $\mathcal{M}_1$ and $\mathcal{M}_2$ is that, in the latter, observations $\{e_i\}_{1\leq i\leq T}$ are not i.i.d.. The value of the i-th recorded EPSC is a function of the number of available and released vesicles n_i and k_i, which in turn depend on their previous values and on the ISI Δt_i. This has two main consequences. Firstly, using the same approximation as in (4.18) would lead to a biased estimate of the identifiability domain. Secondly, the classical definition of the BIC (4.15) should not be used since observations are correlated. Rather, we use Eq. (4.16) to compare the evidence for $\mathcal{M}_1$ and $\mathcal{M}_2$ for a given data set.

Plotting the identifiability domain of a model also allows to investigate how the identifiability depends on the experimental protocol. For model $\mathcal{M}_1$, we already saw that the identifiability domain increases with the number of data points T (see Figure 4.5 (**B**) and (**D**)): a larger data set is more informative and allows for more reliable inference. In this case, T is the only experimental variable, as observations $\{e_i\}_{1\leq i\leq T}$ are i.i.d. On the other hand, the identifiability domain of $\mathcal{M}_2$ will depend not only on the number of data points, but also on the stimulation protocol. We compare the constant stimulation protocol (T data points with a constant inter-spike interval $ISI = 0.05s$) of Figure 4.6 (**A**) with a more realistic stimulation protocol in Figure 4.6 (**B**). In electrophysiological recordings, synaptic transmission is classically studied by stimulating the presynaptic cell with short regular train of spikes at a given frequency, followed by a recovery spike. This protocol is then repeated several times [159, 20, 5]. Such periodic trains are more informative than a constant stimulation protocol, as they allow to probe a broader range of temporal dynamics.

In Figure 4.6 (**B**), we use 20 repetitions of a train of 4 spikes at 20Hz ($ISI = 0.05s$), followed by a recovery spike $0.5s$ later. This protocol entails the same number of data points $T = 100$ as the constant one, but allows to identify STD for a broader range of depression time constants (namely, for $\tau_D < 0.3s$). On the other hand, since there are fewer successive stimulations within a short time interval than in the constant protocol, depression can only be identified when the release probability p is sufficiently high to induce vesicle pool depletion.

4.4.3 Data free model selection

In model-based inference of synaptic parameters, a crucial step related to the estimation of the parameters is model selection, which is usually performed in several steps:

1. Data $\mathcal{D}$ are acquired from a synapse using protocol Ψ;
2. A nested family of $n+1$ possible models $\mathcal{F} = \{\mathcal{M}_0, \mathcal{M}_1, ..., \mathcal{M}_n\}$ is defined;
3. Each of these models is fitted on $\mathcal{D}$ to obtain $n+1$ MLE $\hat{\theta}_0, \hat{\theta}_1, ..., \hat{\theta}_n$;

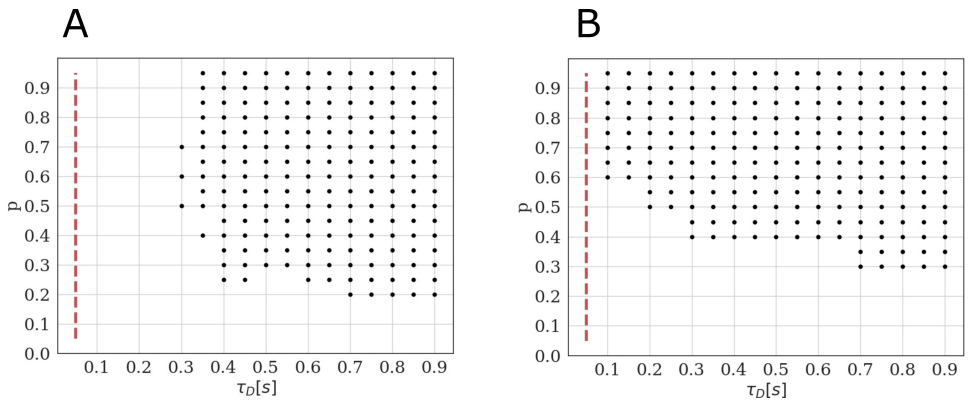


Figure 4.6: Identifiability domain of the binomial model with short-term depression $\mathcal{M}_2$ as a function of p and τ_D. Black dots correspond to the parameters for which the average BIC of 400 data sets results in the correct identifiability of the depressed model. Results obtained for $N = 5$, $q = 1$, $\sigma = 0.2$, and $T = 100$. (**A**) Constant stimulation protocol with an inter-spike interval $ISI = 0.05s$ (red dotted line). (**B**) Stimulation protocol consisting in 20 repetitions of the same spike train: 4 spikes with an inter-spike interval $ISI = 0.05s$ followed by a recovery spike $0.5s$ later.

4. A model selection criterion (Bayes Factor, BIC, AIC...) is computed to quantify and rank the fitness of each model on $\mathcal{D}$;
5. If $\mathcal{M}_i$ is the selected model, then its MLE $\hat{\theta}_i$ is selected as the inference of synaptic parameters.

However, in many studies [20, 5, 26], such a model selection step is not described. In this section, we investigate the possibility, having only access to the inferred values $\hat{\theta}$ of the parameters and to the description of the experimental protocol Ψ, to verify that the model used to infer $\hat{\theta}$ was indeed practically identifiable (i.e. to verify if a simpler model would have given a better fit to the data).

We use the notation $\mathcal{D}$ for a set of data generated from a model $\mathcal{M}$ parametrized with θ^*, and $\hat{\theta}$ the inferred parameters obtained by fitting the parameters θ of model $\mathcal{M}$ on $\mathcal{D}$. If θ^* is within the practical identifiability domain of $\mathcal{M}$ as we defined it, it is then possible to correctly infer it from $\mathcal{D}$, and hence $\hat{\theta} \approx \theta^*$ will also be within the identifiability domain of $\mathcal{M}$. Reciprocally, if $\hat{\theta}$ is not in the identifiability domain of $\mathcal{M}$, then a submodel would have provided a better fit to the data $\mathcal{D}$ than $\mathcal{M}$. Is it thus possible to verify if $\mathcal{M}$ overfits the data simply by verifying if $\hat{\theta}$ is in its identifiability domain, without having access to the data.

This is illustrated in Figure 4.7 (**A**) and (**B**), where $\mathcal{M}_1$ is fitted on data generated from its submodel $\mathcal{M}_0$. For six different values of θ_0^* (**A**), the inferred parameters are out of the identifiability domain of $\mathcal{M}_1$ (**B**), showing that data are indeed better explained by $\mathcal{M}_0$ than by $\mathcal{M}_1$.

First example: application to the data from Katz et al. (1954)

We first apply our *data free* model selection method to the seminal 1954 paper from J. del Castillo and B. Katz [9], in which the quantal nature of neurotransmitter release is identified for the first time. In order to observe mEPSP, they artificially reduced the release probability p by lowering the external calcium concentration. Although the quantal components of postsynaptic potentials are clearly visible and thoroughly

analyzed, it would be interesting to verify, using our proposed model identifiability analysis method, that the binomial model (i.e. a multi-quantal distribution) indeed provides a better fit to the data than a simpler Gaussian model (i.e. a uni-quantal distribution).

Data [69] consist in 328 EPSPs recorded at the neuro-muscular junction (NMJ) of a frog muscle. Fitting the binomial model and running the Expectation-Maximization algorithm on them yields $\hat{N} = 42$, $\hat{p} = 0.013$, $\hat{q} = 0.875mV$, and $\hat{\sigma} = 0.15mV$ (and hence $\frac{\hat{\sigma}}{\hat{q}} \approx 17\%$). For this particular example, we have not only access to the inferred parameters $\hat{\theta}$, but also to the data: it is thus possible to directly compare the BIC of a Gaussian ($\text{BIC}_{\mathcal{M}_0} = 764.95$) and of a binomial ($\text{BIC}_{\mathcal{M}_1} = 470.37$) distributions, which indeed confirms that data are better explained by the binomial quantal model.

However, even without the data, we can verify that the point in the parameter-protocol space specified by $\hat{\theta}$ (the inferred values of the parameters) and Ψ (the number of data points $T = 328$) is indeed within the identifiability domain of the binomial model $\mathcal{M}_1$ compared to $\mathcal{M}_0$ (see Figure 4.7 (**C**)), thus confirming the multi-quantal nature of the recordings.

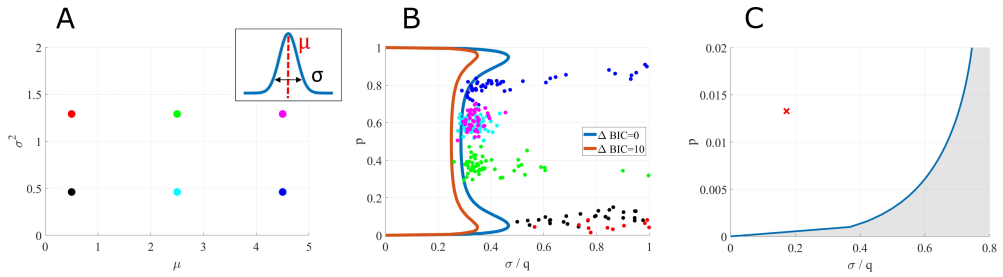


Figure 4.7: (**A**) Data sets were generated from a Gaussian model $\mathcal{M}_0$ (for 6 different means μ and variances σ^2). (**B**) A binomial model $\mathcal{M}_1$ was fitted on them. In each case, inferred parameters (coloured dots) are out of the identifiability domain of $\mathcal{M}_1$. It is thus possible to verify if a model used to fit data was indeed identifiable, without having access to the data and only using inferred parameters. (**C**) Blue line: identifiability domain of the binomial model compared to a Gaussian distribution, for $N = 42$ and $T = 328$, computed from (4.18). The red cross corresponds to the parameters inferred from [9], and is indeed within the identifiability domain.

Second example: application to the data from Barri et al. (2016)

We then apply our method to the results presented in the 2016 paper from A. Barri, Y. Wang, D. Hansel, and G. Mongillo [20], in which the complete binomial model (with STD and STF) is fitted on recordings from layer 5 pyramidal neurons. They use a slight variation of the binomial release model with short term plasticity described by Eq. (4.1) to (4.6), in which the emission probability does not follow a Gaussian, but an inverse Gaussian distribution:

$$p_\theta(e_i|k_i) = \frac{q^{3/2}k_i}{\sqrt{2\pi\sigma^2 e_i^3}} \exp\left(-\frac{q(e_i - qk_i)^2}{2\sigma^2 e_i}\right) \tag{4.19}$$

To verify that the data would not have been better fitted by a simpler model (and hence, that the published estimates of synaptic parameters are reliable), 100 synthetic

data sets were generated from the complete binomial model using the stimulation protocol and the inferred values of the parameters described in [20]:

- 20 repetitions of the same stimulation protocol consisting in 8 presynaptic spikes at 20Hz followed by a recovery spike 500ms later;
- $N^* = 17$, $p^* = 0.27$, $q^* = 0.18mV$, $\sigma^* = 0.06mV$, $\tau_D^* = 202ms$, and $\tau_F^* = 449ms$.

$\mathcal{M}_0$, $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ were then fitted on the generated data. Average values of their respective BIC are presented in Figure 4.8, and confirm the identifiability of the model used in the study.

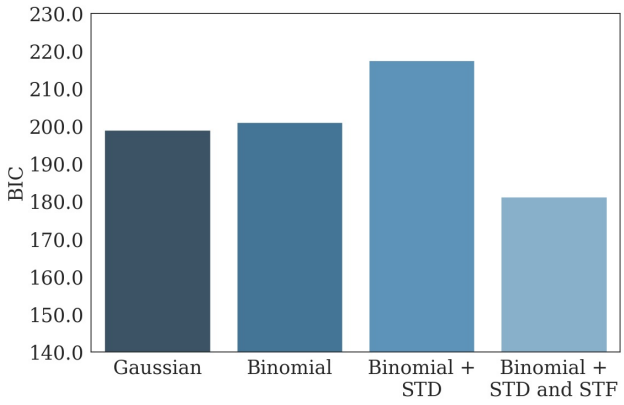


Figure 4.8: Average BIC of $\mathcal{M}_0$, $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ when fitted on 100 independent data sets generated from $\mathcal{M}_3$ parametrized with $N^* = 17$, $p^* = 0.27$, $q^* = 0.18mV$, $\sigma^* = 0.06mV$, $\tau_D^* = 202ms$, and $\tau_F^* = 449ms$. $\mathcal{M}_3$ has the lowest average BIC compared to its submodels, showing that the parameters used to generate the data are indeed within the identifiability domain of $\mathcal{M}_3$. As a consequence, we can infer that $\mathcal{M}_3$ indeed provided the best fit to the data compared to its submodels, and that inferred parameters presented in [20] are reliable. The facilitating nature of the synapse is illustrated by the fact that $\mathcal{M}_2$ (the model with only STD and no STF) has the highest average BIC.

4.5 Discussion

Obtaining an accurate estimate of the parameters of a system from noisy and scarce observations is a crucial problem in neuroscience. Especially, different methods have been proposed for estimating the parameters describing a synapse (namely, its number of independent release sites, their release probability upon the arrival of a presynaptic spike, the quantum of current elicited by one release event, the time constants of depression and facilitation, etc.) Inferring their values allows to analyze the locus of synaptic plasticity and homeostasis; to study possibly synapse-related diseases; and more generally to investigate learning, memory, and neural dynamics, which are underlied by synaptic transmission.

It is usually impossible to measure directly these parameters. However, they can be estimated by fitting a biophysical model of synapse on currents recorded on the post-synaptic side and elicited by experimental stimulation of the presynaptic cell. This approach for estimating the parameters of a system is referred to as model-based

inference. As different competing models may be used to describe the system and explain its output, model-based inference of parameters thus raises the question of what makes a good model.

Prior to any data recording, a required property for competing models is identifiability. Although structural identifiability has been widely studied, no quantitative criterion exists for practical identifiability, which is usually only qualitatively assessed. Here, we propose a definition for the practical identifiability of a model, based on its expected support given the distribution of the data. We define the practical identifiability domain of a statistical model as the set of parameters for which the model is correctly identified as the ground truth compared to a simpler alternative submodel, and we study the identifiability domains of different models of synaptic release. In the process, we propose an extension of the Bayesian Information Criterion (BIC) for models with correlated data. The BIC is a widely used tool for model selection, but it is derived by assuming that the outputs of the system are mutually independent, which is not the case for models of chemical synapse. Finally, we show that our approach allows to perform *data free* model selection, i.e. to verify the identifiability of a model without having access to the data.

The definition of practical identifiability we introduced here differs from the influential contribution of [70] in two ways. Firstly, our definition is data-independent: it does not only hold for a specific set of recorded data $\mathcal{D}$. Indeed, we define practical identifiability as an intrinsic property of the model $\mathcal{M}$ and experimental protocol Ψ. We actually define the *a priori* expected identifiability before a specific $\mathcal{D}$ is recorded, which allows to study how identifiability is affected by different experimental protocols. Secondly, since our definition is derived from a model-selection argument, it does not require to select a possibly arbitrary threshold on the practical identifiability of parameters. Rather, it is defined with respect to a particular submodel. Although the choice of the submodel might itself be arbitrary, we argue that nested models and families naturally arise in commonly used statistical techniques, such as polynomial regression [148], or Generalized Linear Models (GLM) [160]. Especially, the widespread use of phenomenological models in neuroscience [161, 162, 163, 164] makes the use of nested families and submodels relevant.

Another limitation of our approach is its practical implementation. As mentioned, the model evidence $p(\mathcal{D}|\mathcal{M})$, on which our definition is based, is often intractable in practice for complex models, and needs to be estimated. For practical purpose, we used the Bayesian Information Criterion (BIC) to compute the identifiability domains of our different models of synapse. However, we acknowledge that the BIC only provides a valid approximation of the model evidence when the number of samples is sufficiently large. A future step would be to study the robustness of our approach to different computations of the model evidence or to other approximations, such as the Akaike Information Criterion (AIC) [57].

Our identifiability domains are similar to the approach adopted in [73], in which the authors study under which regime of rate fluctuation are the temporal variations of a neuron firing rate correctly identified. Spike trains are generated from a model of spiking neuron with a fluctuating firing rate (complex model); but under a certain value of rate fluctuation, this model becomes indistinguishable from a model of spiking neuron with a constant rate (simple model). Plotting the identifiability of the fluctuating-rate model as a function of the amplitude of rate fluctuation allows them to identify which distribution of inter-spike intervals has the broader identifiability domain (and thus maximizes the efficiency of rate fluctuation transmission).

In model-based inference and parameter estimation, one is often interested in obtaining theoretical bounds on the achievable error performance. Such theoretical bounds allow to assess *a priori* the possibility to correctly infer the parameters. A well-known theoretical result is the Cramér-Rao bound [165, 166], which provides a lower bound on the variance of the parameter estimator. This bound, which depends on the model, its parameters, and the experimental protocol, may actually be too loose in practice, and does not account for the threshold effect described in [167]. In many cases, as the number of data points increases, the estimate error displays a threshold-like transition, from a region of low performance to a region of high performance where the Cramér-Rao bound is attained. Our definition of practical identifiability also discriminates between regions of low information (for small signal-to-noise ratios and sample size) and high accuracy, provides a quantitative criterion to discriminate them, and can be extended to the case of non-i.i.d. data. An interesting future step would be to verify how the boundaries of our proposed identifiability domains compare with the transition threshold described in [167].

An interesting topic would be to study the practical identifiability domain as the number of observations T goes to infinity. In this asymptotic case, practical non-identifiability means that the model cannot be identified, even with an infinite amount of data. We can conjecture that practical identifiability is equivalent to structural identifiability in this asymptotic case, as hinted by Figure 4.5: the identifiability domain increases with T. A future step would be to verify if the practical identifiability domain of a model is included in its structural identifiability domain, and how it behaves when the number of observations T goes to infinity.

We applied our analysis to 4 variants of the binomial model, of increasing complexity: a Gaussian model (i.e. a uni-quantal distribution); a binomial model without short-term plasticity; a binomial model with only short-term depression; and a binomial model with both short-term depression and facilitation. A future step would be to extend our analysis to further generalisations of the binomial model, in order to account for parameters heterogeneity. Especially, the binomial model assumes that the release probability and the quantal amplitude are identical for each release site. It is however possible to hypothesize that there are several pools of vesicles, each having different parameters (for instance a fast depleting pool and a slow depleting pool). There will be regimes in which those sub-pools can be detected and other in which the noise is too high or the experimental protocol not informative enough to identify them, which can be quantified using our definition of identifiability. Another possible generalization of the binomial model is to assume that the postsynaptic response to one vesicle release is not fixed, but follows for instance a Gamma distribution [168] to account for variability in vesicles size and neurotransmitter content.

Model selection is not only a first step in model-based inference of synaptic parameters (as it is necessary to have a reliable estimates of the parameters), but also a tool to study the mechanisms of neurotransmitter release at a chemical synapse. An alternative hypothesis (e.g. *"this synapse shows short-term plasticity"*) can be compared to a null hypothesis (*"this synapse does not show short-term plasticity"*) by computing how well the complex model (i.e. with short-term plasticity) explains the behavior of the synapse compared to the simple model (i.e. without short-term plasticity). Testing models of growing complexity allows to study the nature of the synapse and to identify mechanisms of neurotransmitter release. But the possibility to correctly select the model that corresponds to the true behavior of the synapse will depend on its parameters and on the experimental protocol used to record data:

there are regimes in which the specific features of a model do not appear in the data. Such regimes correspond to the identifiability domain of the model, and studying them allows to draw conclusions on the nature of the synapse.

As stated previously, the problem of inferring parameters from noisy and scarce observations is not restricted to synaptic parameters estimation, but is a crucial question in neuroscience. Our proposed methodology could also be applied to the modelization of single neurons [169, 170, 171, 172], of dynamics of neural populations [173], or of calcium-driven vesicles fusion [174, 175, 176].

On a broader scale, instead of seeing parametric non-identifiability as a statistical problem, we could consider it as a biophysical feature. The total synaptic strength between two cells is a function of both presynaptic (N, p) and postsynaptic (q) parameters. Different combinations of these parameters could lead to the same average postsynaptic response: a presynaptic modification of the number of release sites N can be compensated by an inverse modification of the postsynaptic number of receptors affecting q. This combined effect of presynaptic and postsynaptic plasticity has been shown to enable reliable and flexible learning [137] and homeostatic modulation [32]. More generally, the question of degeneracy, defined as the ability of different elements to perform the same function, could be addressed within the framework of identifiability analysis [177, 178].

Finally, our proposed definition of model identifiability is paving the way towards Optimal Experiment Design (OED) for model selection and parameter inference. The information conveyed by the data about the ground truth model and its parameters depends on the experimental protocol: number of recorded data points, stimulation frequency, etc. The goal of OED is to optimize the experimental protocol in order to maximize the possibility to discriminate between competing models [82, 83] and the precision of the inference of their parameters. An OED for inferring the parameters of a given model maximizes the mutual information between the data and the parameters $\mathcal{I}(\mathcal{D}, \theta)$ [79]. This quantity turns out to be equal to the expected gain in information about θ (defined as the Kullback-Leibler divergence between its prior and its posterior), on which our proposed definition of the informative domain (4.11) is based. Similarly, maximizing the Average Log Bayes Factor (4.13) is equivalent to maximizing the discriminability between the two models $\mathcal{M}$ and $\mathcal{M}'$, and hence finding an OED for model selection. As a thorough theoretical preliminary analysis of the properties of the competing models is a first step prior to model selection and parameter inference [179], we believe that our theoretical contribution to model analysis will contribute to the development of OED techniques for synaptic transmission study.

4.6 Supplementary material

We use the following notations:

- $\mathbf{n} = \{n_i\}_{1 \leq i \leq T}$, $\mathbf{k} = \{k_i\}_{1 \leq i \leq T}$, and $\mathcal{D} = \{e_i\}_{1 \leq i \leq T}$ represent respectively the number of available vesicles at the moment of spike i, the number of vesicles released after spike i (hidden states), and the i-th recorded EPSC (observations);
- θ_l and θ_m are the l-th and m-th elements of the vector θ;
- For simplicity, we use the shorthand $p_\theta(\mathcal{D}) = p(\mathcal{D}|\theta, \mathcal{M})$.

4.6.1 Derivation of the BIC

We define $Q(\theta) = \log p(\mathcal{D}|\theta, \mathcal{M})$, which can be approximated using a second-order Taylor series around the MLE $\hat{\theta}$:

$$Q(\theta) \approx Q(\hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})^T H(\hat{\theta})(\theta - \hat{\theta}) \tag{4.20}$$

where $H(\hat{\theta})$ is the Hessian matrix of Q expressed at $\hat{\theta}$. Using (4.20) in $p(\mathcal{D}|\mathcal{M}) = \int_\theta p(\mathcal{D}|\theta, \mathcal{M})\pi(\theta)d\theta$, and assuming a flat prior for θ given $\mathcal{M}$ yields

$$p(\mathcal{D}|\mathcal{M}) \approx \exp(Q(\hat{\theta}))\pi(\hat{\theta}) \int_\theta \exp\left(\frac{1}{2}(\theta - \hat{\theta})^T H(\hat{\theta})(\theta - \hat{\theta})\right) d\theta$$

Since $H(\hat{\theta})$ is symmetric and negative-definite, the integral can be computed using Laplace's method:

$$\log p(\mathcal{D}|\mathcal{M}) \approx \log p(\mathcal{D}|\hat{\theta}, \mathcal{M}) + \log \pi(\hat{\theta}) + \frac{k_\mathcal{M}}{2}\log(2\pi) - \frac{1}{2}\log(|-H(\hat{\theta})|) \tag{4.21}$$

Assuming that the number T of data points is sufficiently large, and that data in $\mathcal{D}$ are i.i.d., the following simplification can be made:

$$\begin{aligned}[H(\hat{\theta})]_{l,m} = \frac{\partial^2}{\partial\theta_l\partial\theta_m}\log p(\mathcal{D}|\theta, \mathcal{M})\Big|_{\hat{\theta}} &= \frac{\partial^2}{\partial\theta_l\partial\theta_m}\sum_{i=1}^{T}\log p(e_i|\theta, \mathcal{M})\Big|_{\hat{\theta}} \\ &= T\frac{1}{T}\sum_{i=1}^{T}\frac{\partial^2}{\partial\theta_l\partial\theta_m}\log p(e_i|\theta, \mathcal{M})\Big|_{\hat{\theta}} \end{aligned} \tag{4.22}$$

$$\implies [H(\hat{\theta})]_{l,m} \approx -T\mathcal{I}(\hat{\theta})$$

where $\mathcal{I}(\hat{\theta})$ is the Fisher Information Matrix for a single observation expressed at $\hat{\theta}$. (4.21) thus becomes

$$\log p(\mathcal{D}|\mathcal{M}) \approx \log p(\mathcal{D}|\hat{\theta}, \mathcal{M}) + \log \pi(\hat{\theta}) + \frac{k_\mathcal{M}}{2}\log(2\pi) - \frac{k_\mathcal{M}}{2}\log(T) - \frac{1}{2}\log(|\mathcal{I}(\hat{\theta})|)$$

Finally, ignoring the terms that do not depend on T (for T sufficiently large) and multiplying both sides of the equality by -2 leads to $BIC_\mathcal{M}(\mathcal{D}) \approx -2\log p(\mathcal{D}|\mathcal{M})$.

4.6.2 Inferring the values of synaptic parameters using the Expectation-Maximization algorithm

We define the quantity Q as

$$Q(\theta|\theta^t) = \langle \log p_\theta(\mathcal{D}, \mathbf{n}, \mathbf{k})\rangle_{p_{\theta^t}(\mathbf{n},\mathbf{k}|\mathcal{D})}$$

that is to say the expected value of the log likelihood function of θ $\log p_\theta(\mathcal{D}, \mathbf{n}, \mathbf{k})$ with respect to the conditional distribution of the hidden variables $\mathbf{n}$ and $\mathbf{k}$ given the observations $\mathcal{D}$ and the current estimates of the parameters θ^t. The M-step of the algorithm corresponds to the maximization of this quantity with respect to θ:

$$\theta^{t+1} = \arg\max_\theta Q(\theta|\theta^t)$$

which can be realized by computing the null-point of its gradient with respect to θ:

$$\langle \nabla_\theta \log p_\theta(\mathcal{D}, \mathbf{n}, \mathbf{k}) \rangle_{p_{\theta^t}(\mathbf{n},\mathbf{k}|\mathcal{D})} \Big|_{\theta^{t+1}} = 0 \tag{4.23}$$

The derivative of the joint probability with respect to each of the parameters can be analytically computed and then introduced into (4.23). Whenever possible, we use an analytical solution for θ^{t+1}, in order to accelerate each iteration of the algorithm. For any quantity $f(n_i, k_i)$, the average is classically computed as

$$\langle f(n_i, k_i) \rangle_{p_{\Theta^t}(\mathbf{n},\mathbf{k}|\mathcal{D})} = \sum_{\mathbf{n},\mathbf{k}} f(n_i, k_i) p_{\Theta^t}(\mathbf{n}, \mathbf{k}|\mathcal{D})$$

which can be more conveniently computed by using the marginal distributions of vectors $\mathbf{n}$ and $\mathbf{k}$ that share the same values at time i n_i and k_i:

$$\langle f(n_i, k_i) \rangle_{p_{\Theta^t}(\mathbf{n},\mathbf{k}|\mathcal{D})} = \sum_{n_i,k_i} f(n_i, k_i) p_{\Theta^t}(n_i, k_i|\mathcal{D})$$

Similarly, marginal distributions can be computed for any temporal indexes. For instance:

$$\langle n_i - n_{i-1} + k_{i-1} \rangle_{p_{\Theta^t}(\mathbf{n},\mathbf{k}|\mathcal{D})} = \sum_{n_i,n_{i-1},k_{i-1}} (n_i - n_{i-1} + k_{i-1}) p_{\Theta^t}(n_i, n_{i-1}, k_{i-1}|\mathcal{D})$$

The probability $p_{\Theta^t}(n_i, k_i|\mathcal{D})$ can then be computed using the Baum-Welch algorithm. As $(ln(u))' = \frac{u'}{u}$, we obtain the following equations for $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ respectively:

Binomial model without STP

$$q^{t+1} = \frac{\sum_{i=1}^T \langle e_i k_i \rangle}{\sum_{i=1}^T \langle k_i^2 \rangle}$$

$$\sigma^{t+1} = \sqrt{\frac{\sum_{i=1}^T \langle (e_i - q k_i)^2 \rangle}{T}}$$

$$p^{t+1} = \frac{\sum_{i=1}^T \langle k_i \rangle}{NT}$$

Binomial model with STD

$$q^{t+1} = \frac{\sum_{i=1}^{T}\langle e_i k_i\rangle}{\sum_{i=1}^{T}\langle k_i^2\rangle}$$

$$\sigma^{t+1} = \sqrt{\frac{\sum_{i=1}^{T}\langle (e_i - q k_i)^2\rangle}{T}}$$

$$\frac{dQ(\Theta|\Theta^t)}{d\tau_D} = 0 \implies \sum_{i=2}^{T}\frac{\partial I_i}{\partial \tau_D}\left\langle \frac{n_i - n_{i-1} + k_{i-1}}{I_i} - \frac{N - n_i}{1 - I_i}\right\rangle = 0$$

(the sum starts at $i = 2$ since τ_D does not appear in the probability of the first EPSC) with $I_i = 1 - \exp\left(-\frac{\Delta t_i}{\tau_D}\right)$

$$p^{t+1} = \frac{\sum_{i=1}^{T}\langle k_i\rangle}{\sum_{i=1}^{T}\langle n_i\rangle}$$

Binomial model with STD and STF

$$q^{t+1} = \frac{\sum_{i=1}^{T}\langle e_i k_i\rangle}{\sum_{i=1}^{T}\langle k_i^2\rangle}$$

$$\sigma^{t+1} = \sqrt{\frac{\sum_{i=1}^{T}\langle (e_i - q k_i)^2\rangle}{T}}$$

$$\frac{dQ(\Theta|\Theta^t)}{dp} = 0 \implies \sum_{i=1}^{T}\left\langle \frac{\partial u_i}{\partial p}\left(\frac{k_i}{u_i} - \frac{n_i - k_i}{1 - u_i}\right)\right\rangle = 0$$

with

$$\frac{\partial u_i}{\partial p} = 1 + \frac{\partial u_{i-1}}{\partial p}(1-p)\exp\left(-\frac{\Delta t_i}{\tau_F}\right) - u_{i-1}\exp\left(-\frac{\Delta t_i}{\tau_F}\right)$$

and $\frac{\partial u_1}{\partial p} = 1$

$$\frac{dQ(\Theta|\Theta^t)}{d\tau_F} = 0 \implies \sum_{i=1}^{T}\left\langle \frac{\partial u_i}{\partial \tau_F}\left(\frac{k_i}{u_i} - \frac{n_i - k_i}{1 - u_i}\right)\right\rangle = 0$$

with

$$\frac{\partial u_i}{\partial \tau_F} = (1-p)\left(\frac{\partial u_{i-1}}{\partial \tau_F}\exp\left(-\frac{\Delta t_i}{\tau_F}\right) + u_{i-1}\exp\left(-\frac{\Delta t_i}{\tau_F}\right)\frac{\Delta t_i}{\tau_F^2}\right)$$

and $\frac{\partial u_1}{\partial \tau_F} = 0$

$$\frac{dQ(\Theta|\Theta^t)}{d\tau_D} = 0 \implies \sum_{i=2}^{T}\frac{\partial I_i}{\partial \tau_D}\left\langle \frac{n_i - n_{i-1} + k_{i-1}}{I_i} - \frac{N - n_i}{1 - I_i}\right\rangle = 0$$

The EM algorithm is run for different possible values of N; for each of these varying values, a MLE $\hat{\theta}_N$ is computed for the continuous parameters. The value of $\hat{\theta}_N$ yielding the highest likelihood is the global MLE. To minimize the vulnerability to local minima, each run of the EM algorithm needs to be repeated with different initial values θ^0 for the continuous parameters. However, to minimize computation time while plotting the identifiability domain of $\mathcal{M}_2$, the algorithm was only run once with the parameters (including N) initiated at their ground-truth values (i.e. $\theta^0 = \theta^*$).

4.6.3 Baum-Welch algorithm and computation of the Hessian matrix

The elements of the Hessian matrix $H(\theta)$ are the second derivatives of the log-likelihood

$$[H(\theta)]_{l,m} = \frac{\partial^2}{\partial\theta_l \partial\theta_m} \log p_\theta(\mathcal{D})$$

We define the following elements:

- The transition matrix A:

$$A(n', k', n, k, i) = p_\theta(n_i = n, k_i = k | n_{i-1} = n', k_{i-1} = k')$$

- The initial state distribution:

$$\pi_{n,k} = p_\theta(n_1 = n, k_1 = k)$$

- The observation matrix:

$$b_{n,k}(i) = p_\theta(e_i | n_i = n, k_i = k)$$

- The forward probability $\alpha_{n,k}(i) = p_\theta(e_1, \ldots, e_i, n_i = n, k_i = k)$ which is defined recursively:
  1. $\alpha_{n,k}(1) = \pi_{n,k} b_{n,k}(1)$
  2. $\alpha_{n,k}(i+1) = b_{n,k}(i+1) \sum_{n',k'} \alpha_{n',k'}(i) A(n', k', n, k, i+1)$

Finally:

$$p_\theta(\mathcal{D}) = \sum_{n,k} \alpha_{n,k}(T)$$

$[H(\theta)]_{l,m}$ is computed as follows:

$$\frac{\partial^2}{\partial\theta_l \theta_m} \log p_\theta(\mathcal{D}) = -\frac{1}{p_\theta(\mathcal{D})^2} \frac{\partial p_\theta(\mathcal{D})}{\partial\theta_l} \frac{\partial p_\theta(\mathcal{D})}{\partial\theta_m} + \frac{1}{p_\theta(\mathcal{D})} \sum_{n,k} \frac{\partial^2}{\partial\theta_l \theta_m} \alpha_{n,k}(T)$$

with

$$\frac{\partial p_\theta(\mathcal{D})}{\partial\theta_l} = \sum_{n,k} \frac{\partial \alpha_{n,k}(T)}{\partial\theta_l}$$

and where the derivatives of the α parameters from the Baum-Welch algorithm are computed recursively:

$$\frac{\partial \alpha_{n,k}(i+1)}{\partial \theta_l} = \frac{\partial b_{n,k}(i+1)}{\partial \theta_l} \sum_{n',k'} \alpha_{n',k'}(i) A(n',k',n,k,i+1)$$
$$+ b_{n,k}(i+1) \sum_{n',k'} \left(\frac{\partial \alpha_{n',k'}(i)}{\partial \theta_l} A(n',k',n,k,i+1) + \alpha_{n',k'}(i) \frac{\partial A(n',k',n,k,i+1)}{\partial \theta_l} \right)$$

and

$$\frac{\partial^2}{\partial \theta_l \theta_m} \alpha_{n,k}(i+1) = \frac{\partial^2}{\partial \theta_l \theta_m} b_{n,k}(i+1) \sum_{n',k'} \alpha_{n',k'}(i) A(n',k',n,k,i+1)$$
$$+ \frac{\partial b_{n,k}(i+1)}{\partial \theta_l} \sum_{n',k'} \left[\frac{\partial \alpha_{n',k'}(i)}{\theta_m} A(n',k',n,k,i+1) + \alpha_{n',k'}(i) \frac{\partial A(n',k',n,k,i+1)}{\theta_m} \right]$$
$$+ \frac{\partial b_{n,k}(i+1)}{\partial \theta_m} \sum_{n',k'} \left[\frac{\partial \alpha_{n',k'}(i)}{\theta_l} A(n',k',n,k,i+1) + \alpha_{n',k'}(i) \frac{\partial A(n',k',n,k,i+1)}{\theta_l} \right]$$
$$+ b_{n,k}(i+1) \sum_{n',k'} \left[\frac{\partial^2 \alpha_{n',k'}(i)}{\partial \theta_l \partial \theta_m} A(n',k',n,k,i+1) + \frac{\partial \alpha_{n',k'}(i)}{\partial \theta_l} \frac{\partial A(n',k',n,k,i+1)}{\partial \theta_m} \right.$$
$$\left. + \frac{\partial \alpha_{n',k'}(i)}{\partial \theta_m} \frac{\partial A(n',k',n,k,i+1)}{\partial \theta_l} + \alpha_{n',k'}(i) \frac{\partial^2}{\partial \theta_k \partial \theta_m} A(n',k',n,k,i+1) \right]$$

In order to compute the derivative of these quantities with respect to N, which is a discrete parameter, binomial distributions were approximated by a normal distribution having the same mean and variance.

4.7 Data Availability Statement

`MATLAB` and `Python` files are available in the following GitHub repository: `https://github.com/camillegontier/identifiability_binomial`

4.8 Funding

The work presented in this paper was supported by the Swiss National Science Foundation under grant number 31003A 175644 entitled Bayesian Synapse.

4.9 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4.10 Acknowledgments

We thank Simone Carlo Surace and Mihai A. Petrovici for the fruitful discussions. Figures were created using BioRender. Calculations were performed on UBELIX (`http://www.id.unibe.ch/hpc`), the HPC cluster at the University of Bern.

Chapter 5

Efficient Sampling-Based Bayesian Active Learning

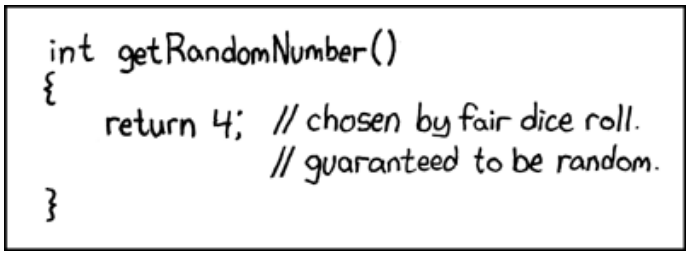


Figure 5.1: On fair sampling [1].

This chapter includes the paper *Efficient Sampling-Based Bayesian Active Learning* [180], available as an ArXiV preprint. Results have been presented during a poster presentation at the Bernstein conference 2020[2] and at the Cosyne conference 2022[3]. The format was adapted to fit that of the thesis and the references have been included in the main bibliography.

Author contributions

Camille Gontier and Jean-Pascal Pfister wrote the paper and derived the equations. Camille Gontier performed the numerical simulations and made the figures. Camille Gontier and Simone Carlo Surace wrote the `Julia` code. Igor Delvendahl and Camille Gontier performed the recordings. All authors contributed to the article and approved the submitted version.

[1] `https://xkcd.com/221/`, use with authorization.

[2] C. Gontier and J.-P. Pfister. Fast and online inference of synaptic parameters. 16$^{\text{th}}$ Bernstein Conference, Berlin, Germany. DOI: `https://doi.org/10.12751/nncn.bc2020.0029`

[3] C. Gontier, S. Surace, J.-P. Pfister. Bayesian active learning for closed-loop synaptic characterization. Cosyne 2022, Lisbon, Portugal.

5.1 Abstract

Bayesian Active Learning (BAL) is an efficient framework for learning the parameters of a model, in which input stimuli are selected to maximize the mutual information between the observations and the unknown parameters. However, its applicability to real experiments is limited by two main drawbacks. Firstly, it requires performing high-dimensional integrations and optimizations in real time: current methods are either too time consuming, or only applicable to specific models. Secondly, it only optimizes for the next stimulus input (an approach referred to as a myopic design), disregarding all future observations in the experiment. To address these issues, we propose an Efficient Sampling-Based Bayesian Active Learning (ESB-BAL) framework, which is efficient enough to be used in real-time biological experiments and to go beyond myopic approaches. We apply our method to the problem of estimating the parameters of a chemical synapse from its postsynaptic responses evoked by presynaptic action potentials. Synaptic parameter inference is a challenging application of BAL, as computation needs to be faster than the smallest inter-stimulation interval, which can be on the order of a few milliseconds. Using synthetic data and synaptic whole-cell patch clamp recordings in cerebellar brain slices, we show that our method has the potential to significantly improve the precision of model-based inferences, thereby paving the way towards more systematic and automated experimental designs.

5.2 Introduction

In machine learning, a central problem is that of inferring the parameters θ of a model $\mathcal{M}$. For instance, in supervised learning, one may want to learn the parameters of a Deep Neural Network (DNN) so as to minimize the difference between its output and training labels; in biology, the parameters of a system can be studied by fitting a biophysical model to recorded observations. In most cases, these parameters can be neither directly measured nor analytically computed, but can be inferred using the recorded outputs of the system $\mathbf{x}$ as a response to input stimuli Ψ. By computing the likelihood of the outputs given the inputs and the parameters $p(\mathbf{x}|\Psi, \theta)$, it is possible to obtain either a point-based estimate of the parameters $\arg\max_\theta p(\mathbf{x}|\Psi, \theta)$ [20], or to use the Metropolis-Hastings (MH) algorithm to compute their full posterior distribution $p(\theta|\mathbf{x}, \Psi) \propto p(\mathbf{x}|\Psi, \theta)$ [5].

However, the accuracy of these estimates critically depends on the pair $(\Psi, \mathbf{x})$, and especially on how the input stimuli are chosen. For instance, training a DNN on non-i.i.d. training examples (i.e. blocked training) will lead to catastrophic forgetting [181]. In biology, most experiments still rely on pre-defined and non-adaptive protocols Ψ, which may not yield sufficient information about the true parameters of the studied system. As a consequence, those experiments require more observations, which increases their cost, time, and need for subjects. An efficient framework to alleviate this issue is called Bayesian Active Learning (BAL). Knowing the current estimate of the parameters, the experimental protocol (i.e. the next input) can be optimized on the fly so as to maximize the mutual information between the recordings and the parameters (Figure 5.2 (a)). BAL is a branch of Optimal Experiment Design (OED) theory [182, 80, 183]. It has already been used in neuroscience to infer the parameters of a Generalized Linear Model (GLM) [29], the receptive field of a neuron [184], or the nonlinearity in a linear-nonlinear-Poisson (LNP) encoding model [185]. Implementing BAL for biological settings can be challenging, especially for

real-time applications: it requires computing an update of the posterior distribution of parameters after each time step, and using it to compute the expected information gain from future experiments, which involves solving an optimization problem over a possibly high-dimensional stimulus space.

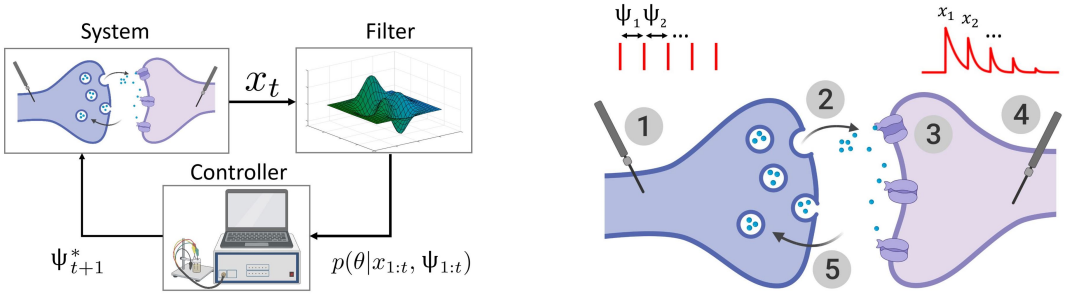
Our main contribution is to provide a general framework for online active learning, called Efficient Sampling-Based Bayesian Active Learning (ESB-BAL). Our novelty is to use particle filtering, which is a highly versatile filtering method [186], for posterior computation; and to propose a parallel computing implementation [187, 188] for efficient posterior update and information computation. While previous implementations of active learning either relied on time consuming Monte Carlo (MC) methods [79, 81] or were only applicable to special cases, such as linear models or GLM [29], our proposed solution is fast enough to be used in real-time biological experiments and can be applied to any state-space model.

To illustrate our method, we apply it to the problem of inferring the parameters of a chemical synapse with Short-Term Depression (STD) [9]. The accuracy of these estimates critically depends on the presynaptic stimulation times: if Inter-Stimulation Intervals (ISIs) are longer than the depression time constant, STD will not be precisely quantified. But if the stimulation frequency is too high, the pool of presynaptic vesicles will be depleted, leading to poor parameter estimates [134, 189]. Synaptic characterization is thus a relevant example application for ESB-BAL, but it is also a challenging one: computation needs to be faster than the typical ISI, which can be shorter than a few milliseconds. Using synthetic data, we show that our method allows to significantly reduce the uncertainty of the estimate in comparison to classically used non-adaptive stimulation protocols. We also show that the rate of information gain (in bit/s) of the whole experiment can be optimized by adding a penalty term for longer ISIs. Lastly, we extend active learning to non-myopic designs. Using recordings from acute brain slices from mice, we show that our framework is sufficiently efficient for optimizing not only the immediate next stimulus, but rather the future stimuli in the experiment.

5.3 Bayesian active learning

When using active learning in sequential experiments, three key elements need to be defined (Figure 5.2): (1) The **system** to be studied: its parameters θ can be inferred from its observed responses to a set of input stimuli Ψ. Given the stochastic nature of most systems studied in biology, the random variable $X_{1:t}$ corresponding to the observations can take various values $x_{1:t}$ according to a distribution $p(x_{1:t}|\Psi, \theta)$ (see Section 5.6); (2) A **filter** that computes the posterior distribution of the parameters given the previous inputs and observations $p(\theta|x_{1:t}, \Psi)$: it is updated after each new observation (see Section 5.4); (3) A **controller** that computes the next input stimuli so as to maximize a certain utility function (see Section 5.5). The three following sections will describe each of these elements in detail.

In general, the utility of a given protocol Ψ can be expressed as the mutual information between the parameter random variable Θ and the response random variable $X_{1:t}$, i.e. $I_\Psi(\Theta; X_{1:t})$ (see [79, 78] for a detailed discussion). For instance, in synaptic characterization, Ψ corresponds to a set of T stimulation times $\psi_{1:T}$ and observations correspond to recorded excitatory postsynaptic currents (EPSCs) $x_{1:T}$. In case of successive experiments [184], the mutual information between the parameters and the next observation x_{t+1} conditioned on the experiment history $h_t = (x_{1:t}, \psi_{1:t})$



(a) (b)

Figure 5.2: (a) Bayesian active learning applied to synaptic characterization. At each time step, the response of the synapse (**system**) to artificial stimulation is recorded. This observation x_t is used by the **filter** to compute the posterior distribution of parameters $p(\theta|x_{1:t}, \psi_{1:t})$. The **controller** then computes the next stimulation time ψ^*_{t+1} to maximize the expected gain of information. (b) Model of binomial synapse with STD. (1): the presynaptic axon is stimulated. (2): After spike t, k_t vesicles (out of the n_t available ones) release their neurotransmitter with a probability p. (3): A single release event triggers a quantal response q. (4): the total recorded postsynaptic response x_t is the sum of the effects of the k_t release events. (5): After releasing, vesicles are replenished with a time constant τ_D, which determines short-term depression.

is:

$$I^{\psi_{t+1}}_{h_t}(\Theta; X_{t+1}) = H_{h_t}(\Theta) - H^{\psi_{t+1}}_{h_t}(\Theta|X_{t+1}) \tag{5.1}$$

where $H_{h_t}(\Theta)$ is the entropy of the distribution of Θ at time t; and

$$H^{\psi_{t+1}}_{h_t}(\Theta|X_{t+1}) = \int dx_{t+1} p(x_{t+1}|h_t, \psi_{t+1}) H^{\psi_{t+1}}_{h_t}(\Theta|X_{t+1} = x_{t+1})$$

is the conditional entropy of the distribution of Θ at time $t+1$. The latter depends on the future observation x_{t+1}, which is unknown. We thus take its average over x_{t+1}; as the predictive distribution depends on the unknown parameters, we also have to take an average over θ, using the current posterior distribution $p(\theta|h_t)$ at time t: $p(x_{t+1}|h_t, \psi_{t+1}) = \int d\theta p(x_{t+1}|h_t, \psi_{t+1}, \theta) p(\theta|h_t)$ [29]. The goal of Bayesian active learning is to select the next stimulation to maximize the mutual information between the parameters and all future observations:

$$\psi^*_{t+1} = \underset{\psi_{t+1} \in \mathcal{S}_{t+1}}{\arg\max} \max_n \max_{\psi_{t+1:t+n}} I^{\psi_{t+1:t+n}}_{h_t}(\Theta; X_{t+1:t+n}) \tag{5.2}$$

with $\psi_{t+1:t+n} \in \mathcal{S}_{t+1:t+n}$ where $\mathcal{S}_{t+1:t+n}$ is the set of possible protocols for the next stimulations. Optimizing all future inputs is an intractable problem (especially for online applications), since the algorithmic complexity scales exponentially with the number of observations. For this reason, BAL only optimizes for the next stimulus (an approach referred to as a *myopic* design) (see Figure 5.2):

$$\psi^*_{t+1} = \underset{\psi_{t+1} \in \mathcal{S}_{t+1}}{\arg\max} I^{\psi_{t+1}}_{h_t}(\Theta; X_{t+1}) \tag{5.3}$$

Different methods have been proposed to compute Eq. 5.3. Monte Carlo (MC) methods [79] or a variational approach [81] can be employed, but they usually require long computation times that can be impractical if the time between successive experiments is short. Closed-form solutions can be computed for some special cases, such as linear models of GLM [29].

5.4 The filter: online computation of the posterior distributions of parameters

To be applicable for online experiments, the filtering block (which will compute the posterior distribution of parameters $p(\theta|h_t)$) needs to satisfy two requirements: (1) It must be sufficiently versatile to be applied to different systems and models; (2) It must be online (i.e. its algorithmic complexity should not increase with the number of observations) [3]. A promising solution is particle filtering [49], and especially the Nested Particle Filter (NPF) [186]. This algorithm is asymptotically exact and purely recursive, and thus allows to directly estimate the parameters as recordings are acquired. The NPF relies on two nested layers of particles to approximate the posterior distributions of both the static parameters of the model and of its hidden states. A first outer filter with M_{out} particles is used to compute the posterior distribution of parameters $p(\theta|h_t)$, and for each of these particles, an inner filter with M_{in} particles is used to estimate the corresponding hidden states z_t (so that the total number of particles in the system is $M_{\text{out}} \times M_{\text{in}}$). After each new observation, these particles are resampled based on their respective likelihoods (Figure 5.6).

The NPF was originally proposed for standard HMMs. Here, we extend it to the more general class of Input-Output Hidden Markov Models (IO-HMMs, also called GLM-HMMs in neuroscience [190]), in which state transition probability at time t depends on an external input ψ_t. For instance, state transition in our model of synapse is not stationary, but depends on the ISI ψ_t (see Section 5.6). The filter (Algorithm 2, Section 5.9.1) relies on the following approximation to recursively compute the likelihood of each particle. Once the observation x_t has been recorded, the likelihood of particle θ_t^i (with i in $1 \dots M_{\text{out}}$) depends on

$$p(\theta_t^i|x_{1:t}) \propto p(x_t|x_{1:t-1}, \theta_t^i, \psi_t) p(\theta_t^i|x_{1:t-1}) \tag{5.4}$$

with

$$p(x_t|x_{1:t-1}, \theta_t^i, \psi_t) = \sum_{z_{t-1:t}} p(x_t|z_t, \theta_t^i) p(z_t|z_{t-1}, \theta_t^i, \psi_t)\ p(z_{t-1}|x_{1:t-1}, \theta_t^i) \tag{5.5}$$

If the variance of the jittering kernel κ (which mutates the samples to avoid particles degeneracy and local solutions, see Section 3.2 in [186] for a detailed discussion) is sufficiently small, and hence if $\theta_t^i \approx \theta_{t-1}^i$, the approximation $p(z_{t-1}|x_{1:t-1}, \theta_t^i) \approx p(z_{t-1}|x_{1:t-1}, \theta_{t-1}^i)$ allows to recursively compute Eq. 5.4. In practice, the different terms in Eq. 5.5 are computed as such: $p(x_t|z_t, \theta_t^i)$ corresponds to the *Likelihood* step of Algorithm 2; $p(z_t|z_{t-1}, \theta_t^i, \psi_t)$ corresponds to the *Propagation* step; and $p(z_{t-1}|x_{1:t-1}, \theta_t^i)$ corresponds to the distribution of hidden states at time $t-1$ (see Section 3.4 in [186] for a detailed explanation).

Contrary to previous methods for fast posterior computation that were only applicable to specific models [29], our filter can be applied to a broad range of state-space

dynamical systems, including non-stationary and input-dependent ones. Moreover, it does not require to approximate the posterior as a Gaussian nor require a time consuming (and possibly unstable) numerical optimization step, while being highly parallelizable and efficient [187, 188].

5.5 The controller: computation of the optimal next stimulation time

The objective of experiment design optimization is to minimize the uncertainty of the estimates (classically quantified using the entropy) while reducing the cost of experimentation (defined as the number of required trials, samples, or observations). The optimal next stimulus ψ^*_{t+1} that will maximize the mutual information (i.e. minimize the uncertainty about θ as measured by the entropy) can be written from Eq. 5.1 and 5.3 as

$$\psi^*_{t+1} = \underset{\psi_{t+1}\in\mathcal{S}_{t+1}}{\arg\min} \int d\theta p(\theta|h_t) \int dx_{t+1} p(x_{t+1}|h_t, \psi_{t+1}, \theta) H_{h_t}^{\psi_{t+1}}(\Theta|X_{t+1} = x_{t+1}) \quad (5.6)$$

Eq. 5.6 requires to compute two (possibly high-dimensional) integrals over θ and x_{t+1}, for which closed-form expressions only exist for specific models. To avoid long MC simulations, we propose to use mean-field computations and to replace integrals by point-based approximations. Firstly, instead of computing the full expectation over $p(\theta|h_t)$, we set θ to its maximum a posteriori (MAP) value $\hat{\theta}_t = \arg\max_\theta p(\theta|h_t)$. Eq. 5.6 thus becomes

$$\psi^*_{t+1} \approx \underset{\psi_{t+1}\in\mathcal{S}_{t+1}}{\arg\min} \int dx_{t+1}\, p(x_{t+1}|h_t, \psi_{t+1}, \hat{\theta}_t) H_{h_t}^{\psi_{t+1}}(\Theta|X_{t+1} = x_{t+1})$$

Depending on the nature of the studied system and on the time constraints of the experiment, different estimators can also be used, such as e.g. $\hat{\theta}_t = \frac{1}{M_{\text{out}}} \sum_{i=1}^{M_{\text{out}}} \theta_t^i$. Secondly, instead of computing the full expectation over the future observation, we set x_{t+1} to its expected value; Eq. 5.6 thus becomes

$$\psi^*_{t+1} \approx \underset{\psi_{t+1}\in\mathcal{S}_{t+1}}{\arg\min}\ H_{h_t}^{\psi_{t+1}}(\Theta|X_{t+1} = \mathbb{E}(X_{t+1}|h_t, \psi_{t+1}, \hat{\theta}_t))$$

In the general case, $\mathbb{E}(X_{t+1}|h_t, \psi_{t+1}, \hat{\theta}_t)$ can be computed using Bayesian Quadrature [191]. More specifically, for our model of a chemical synapse, an analytical formulation for the expected value $\mathbb{E}(X_{t+1}|\psi_{1:t+1}, \hat{\theta}_t)$ can be efficiently derived using mean-field approximations (see Section 5.9.2). For each candidate ψ_{t+1} in a given finite set $\mathcal{S}_{t+1}$, the entropy $H_{h_t}^{\psi_{t+1}}(\Theta|X_{t+1} = \mathbb{E}(X_{t+1}|h_t, \psi_{t+1}, \hat{\theta}_t))$ can be computed using Algorithm 2. Similarly as in [190], the entropy of the posterior distribution of Θ is approximated as $\frac{1}{2}\log|2\pi e\Sigma_t|$, where Σ_t is the covariance matrix of the particles $\{\theta_t^i\}_{1\leq i\leq M_{\text{out}}}$.

5.6 The system: a binomial model of neurotransmitter release

A classically used model to describe the release of neurotransmitters at chemical synapses is called the binomial model [12, 20, 5, 134, 47, 42, 17]. Under this model, a synapse is described as a Input-Output Hidden Markov Model (IO-HMM) with the following parameters (units are given in square brackets, see also Figure 5.2 (b)): N

Algorithm 1: Computation of the optimal next stimulation time for synaptic characterization

set $\hat{\theta}_t = \arg\max_\theta p(\theta|h_t)$ (MAP values from the current posterior estimation);
Input: $\mathcal{S}_{t+1}$ (set of candidates ψ_{t+1});
for ψ_{t+1} *in* $\mathcal{S}_{t+1}$ **do**
  Compute $\mathbb{E}(X_{t+1}|\psi_{1:t+1}, \hat{\theta}_t)$ using Eq. 5.10;
  Compute $H_{h_t}^{\psi_{t+1}}(\Theta|X_{t+1} = \mathbb{E}(X_{t+1}|\psi_{1:t+1}, \hat{\theta}_t))$ using Algorithm 2;
end
$\psi^*_{t+1} = \arg\min_{\psi_{t+1}\in\mathcal{S}_{t+1}} H_{h_t}^{\psi_{t+1}}(\Theta|X_{t+1} = \mathbb{E}(X_{t+1}|\psi_{1:t+1}, \hat{\theta}_t))$

(the number of independent release sites [-]); p (their release probability [-]); σ (the recording noise [A]); q (the quantum of current elicited by one release event [A]); τ_D (the time constant of synaptic vesicle replenishment [s]). The variables n_t, k_t, and x_t represent, respectively, the number of available vesicles at the moment of spike t, the number of vesicles released after spike t, and the t-th recorded EPSC. For simplicity, we use the notation $p_\theta(\cdot) = p(\cdot|\theta)$ with $\theta = [N, p, q, \sigma, \tau_D]$. The probability of recording $x_{1:T}$ is computed as the marginal of the joint distribution of the observations $x_{1:T}$ and the hidden variables $n_{1:T}$ and $k_{1:T}$:

$$p_\theta(x_{1:T}, n_{1:T}, k_{1:T}) = p_\theta(x_1|k_1)p_\theta(k_1|n_1)p_\theta(n_1)\prod_{t=2}^{T} p_\theta(x_t|k_t)p_\theta(k_t|n_t)p_\theta(n_t|n_{t-1}, k_{t-1}, \psi_t) \tag{5.7}$$

where $p_\theta(x_t|k_t) = \mathcal{N}(x_t; qk_t, \sigma^2)$ is the emission probability, i.e. the probability to record x_t knowing that k_t vesicles released neurotransmitter; $p_\theta(k_t|n_t)$ is the binomial distribution and represents the probability that, given n_t available vesicles, k_t of them will indeed release neurotransmitter: $p_\theta(k_t|n_t) = \begin{pmatrix} n_t \\ k_t \end{pmatrix} p^{k_t}(1-p)^{n_t-k_t}$; and $p_\theta(n_t|n_{t-1}, k_{t-1}, \psi_t)$ represents the process of vesicle replenishment. During the time interval ψ_t, each empty vesicle can refill with a probability $I_t(\psi_t) = 1 - \exp\left(-\frac{\psi_t}{\tau_D}\right)$ such that the transition probability $p_\theta(n_t|n_{t-1}, k_{t-1}, \psi_t)$ is given by:

$$p_\theta(n_t|n_{t-1}, k_{t-1}, \psi_t) = \begin{pmatrix} N - n_{t-1} + k_{t-1} \\ n_t - n_{t-1} + k_{t-1} \end{pmatrix} I_t(\psi_t)^{n_t - n_{t-1} + k_{t-1}}(1 - I_t(\psi_t))^{N-n_t} \tag{5.8}$$

Eqs. 5.7 to 5.8 define the observation model of the studied system (see Figure 5.2), i.e. the probability of an observation x_t given a vector of stimuli $\psi_{1:t}$ and a vector of parameters θ.

5.7 Results

5.7.1 First setting: reducing the uncertainty of estimates for a given number of observations

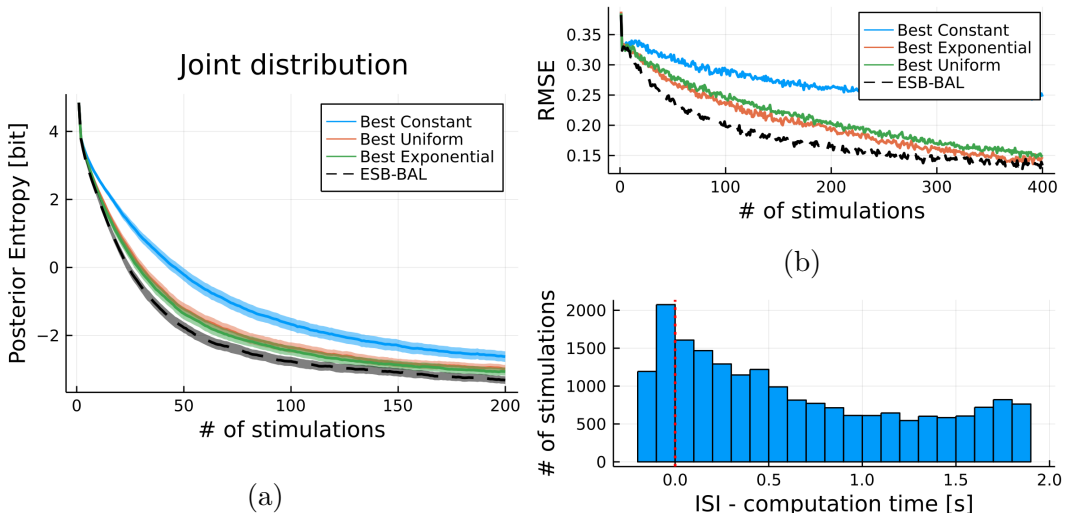
To illustrate our ESB-BAL framework, we apply it to the problem of estimating the parameters of a chemical synapse (Algorithm 1). From the experimentalist point of view, a highly relevant question is how to optimize the stimulation protocol such that the measured EPSCs are most informative about synaptic parameters. Previous

studies showed that some stimulation protocols are more informative than others, but ignored the temporal correlations of the number of readily-releasable vesicles [159] or did not compute which protocol would be most informative [20].

Results for a simulated experiment with ground-truth parameters $N^* = 7$, $p^* = 0.6$, $q^* = 1$ pA, $\sigma^* = 0.2$ pA, and $\tau_D^* = 0.25$s (i.e. the same set of parameters θ^* used in [5]) are displayed in Figure 5.3 (a). Here, we compare ESB-BAL to 3 deterministic protocols: in the *Constant* protocol, the synapse is probed at a constant frequency, i.e. $\Delta_t = \text{cst}$; in the *Uniform* protocol, ISIs are uniformly drawn from a set $\mathcal{S}$ of candidates ψ_t consisting of equidistantly separated values ranging from $\Delta_t^{\min} = 0.005$s (i.e. one order of magnitude shorter than the shortest ISI used in [20]) to $\Delta_t^{\max}$, i.e. $\psi_t \sim \mathcal{U}([0.005, \Delta_t^{\max}])$; finally, in the *Exponential* protocol, ISIs are drawn from an exponential distribution with mean τ. The efficiency of these deterministic protocols will depend on their respective parametrizations. To conservatively assess ESB-BAL, we optimize the values of Δ_t, $\Delta_t^{\max}$, and τ so that the *Constant*, *Uniform*, and *Exponential* protocols have the best possible performance for the used ground-truth parameters θ^*. Figure 5.7 shows the average final entropy decrease (i.e. the information gain) after 200 observations using the *Constant* (top), *Uniform* (middle), or *Exponential* (bottom) protocol, for different values of their hyperparameters. These deterministic protocols (with their optimal respective parametrizations) are then compared to ESB-BAL.

For the different protocols, the average (over 100 independent repetitions) joint differential entropy of the posterior distribution of parameters is plotted as a function of the number of observations. ESB-BAL allows to significantly reduce the uncertainty (as measured by the entropy) of the parameter estimates for a given number of observations. It should be noted that it is compared to deterministic protocols whose respective hyperparameters have been optimized offline, knowing the value of θ^*. In real physiology experiments, classical protocols are non-adaptative and are defined using (possibly sub-optimal) default parameters. In contrast, in active learning the protocol is optimized on the fly as data are recorded, and its performance will not depend on a prior parametrization. ESB-BAL thus outperforms the best possible *Constant*, *Uniform*, and *Exponential* protocols.

We also verify that ESB-BAL does not lead to biased estimates of θ, as its average Root-Mean-Square Error (RMSE) outperforms that of other protocols (Figure 5.3 (b)); and that it is sufficiently fast for online applications, as computation time exceeds the ISI in only a small proportion of cases (Figure 5.3 (c)). Similar results can be observed for different sets of ground-truth parameters θ^* (Figure 5.8) or when only optimizing for the entropy of a specific parameter (Figure 5.9). Finally, in Figure 5.10, ESB-BAL (black dashed line) is compared to exact active learning (gray dashed line), in which Eq. 5.6 is computed exactly using MC samples. Samples to compute the expectation over θ are drawn from $p(\theta|h_t)$, whereas samples used to compute the expectation over x_{t+1} are drawn from a normal distribution with mean $\mathbb{E}(X_{t+1})$ (Eq. 5.10) and variance $\text{Var}(X_{t+1})$ (Eq. 5.11). This shows that the approximations used in Algorithm 1 to make active learning online have only a small effect on performance.



(a) (b) (c)

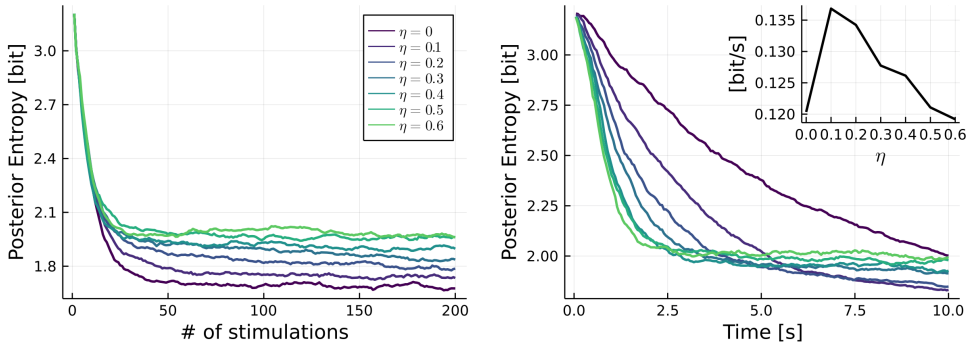
Figure 5.3: **First setting: reducing the uncertainty of estimates for a given number of observations.** (a) Entropy of the posterior distribution of θ vs. number of observations for different stimulation protocols. Synthetic data were generated from a model of synapse with ground truth parameters $N^* = 7$, $p^* = 0.6$, $q^* = 1$ A, $\sigma^* = 0.2$ A, and $\tau_D^* = 0.25$s [5]. Traces show average over 100 independent repetitions. Shaded area: standard error of the mean. (b) RMSE for the same simulations. (c) Histogram of the differences between the ISI and the corresponding computation time for the ESB-BAL simulations.

5.7.2 Second setting: reducing the uncertainty of estimates for a given experiment time

Active learning allows, for a given number of observations, to improve the reliability of the estimated parameters. However, in its classical implementation, only the next stimulus input is optimized, disregarding all future observations in the experiment. This myopic approach is thus sub-optimal. Moreover, neurophysiology experiments are not only constrained by the number of observations, but also by the total time of the experiment. Since cell viability and recording stability may become limiting during an experiment, the total time of an experimental protocol $\sum_{t=1}^{T} \psi_t$ also needs to be accounted for. Here, to account for the total time of the experiment, and to globally optimize the information gain per unit of time, we propose to modify the classical formulation of active learning (Eq. 5.6) by adding a penalty term for longer ISIs:

$$\psi_{t+1}^{*(\eta)} = \underset{\psi_{t+1}}{\arg\min} \left\{ \eta\psi_{t+1} + \int d\theta p(\theta|h_t) \int dx_{t+1} p(x_{t+1}|\psi_{1:t+1}, \theta) H_{h_t}^{\psi_{t+1}}(\Theta|x_{t+1}) \right\} \tag{5.9}$$

The effect of the penalty weight η on the entropy of the posterior distribution of τ_D is displayed in Figure 5.4 (a). As expected, adding a penalty term to Eq. 5.6 reduces the precision of the inferred parameter. The loss of information gain increases with the penalty weight η. However, increasing η also increases the speed of information gain, as seen in Figure 5.4 (b). Depending on the available time for the experiment, it



(a) (b)

Figure 5.4: **Second setting: reducing the uncertainty of estimates for a given experiment time** (effect of penalizing long ISIs on parameter estimates uncertainty and rate of information gain). (a) Entropy of the posterior distribution of τ_D vs. number of observations for different values of η in Eq. 5.9. Same settings as in Figure 5.8. (b) Same results, but displayed as a function of time. Inset: slope of the entropy vs. time curves (i.e. information rate) vs. η after 10 seconds.

is thus possible to tune η so as to find a trade-off between long-term precision (Figure 5.4 (a)) and information rate (Figure 5.4 (b)).

5.7.3 Third setting: batch optimization and application to neural recordings

To reduce computational complexity, classical implementations of sequential experiment design usually only optimize for the immediate next observation. However, it might be critical for some systems to optimize not only the next stimulus, but rather the next n stimuli of the experiment altogether (see Eq. 5.2) [183, 192]. Synaptic characterization is a telling example: indeed, STD can only be observed for specifically organized batches of stimulation times. When probing the presynaptic cell, neuroscientists usually use repetitions of a spike train (Figure 5.5 (b)) consisting of a tetanic stimulation phase (sustained high-frequency stimulation used to deplete the presynaptic vesicles, between 0.1s and 0.3s in Figure 5.5 (b)) followed by recovery spikes at increasing ISIs to probe the STD time constant [193]. These spike trains are usually not optimized, and are held constant throughout an entire experiment.

Here, we show that ESB-BAL can be used to extend active learning to non-myopic designs, and to optimize the n next input stimuli. Algorithm 3 (Section 5.9.4), which is a generalization of Algorithm 1, is used to select the next batch of n stimuli $\psi^*_{t+1:t+n}$ in a set of candidate batches $\mathcal{S}_{t+1:t+n}$. Every n observations, $H_{h_t}^{\psi_{t+1:t+n}}(\Theta|X_{t+1:t+n})$ is computed using n iterations of the filter (i.e. Algorithm 2), so as to pick the optimal next batch $\psi^*_{t+1:t+n}$ that minimizes this quantity:

$$\psi^*_{t+1:t+n} = \underset{\psi_{t+1:t+n} \in \mathcal{S}_{t+1:t+n}}{\arg\min} \; H_{h_t}^{\psi_{t+1:t+n}}(\Theta|X_{t+1:t+n})$$

where the next stimulation times in $\mathcal{S}_{t+1:t+n}$ are parametrized with a low-dimensional parametrization so as to span different durations and frequencies for the tetanic phase, and different ISIs between the recovery spikes (see Figure 5.11).

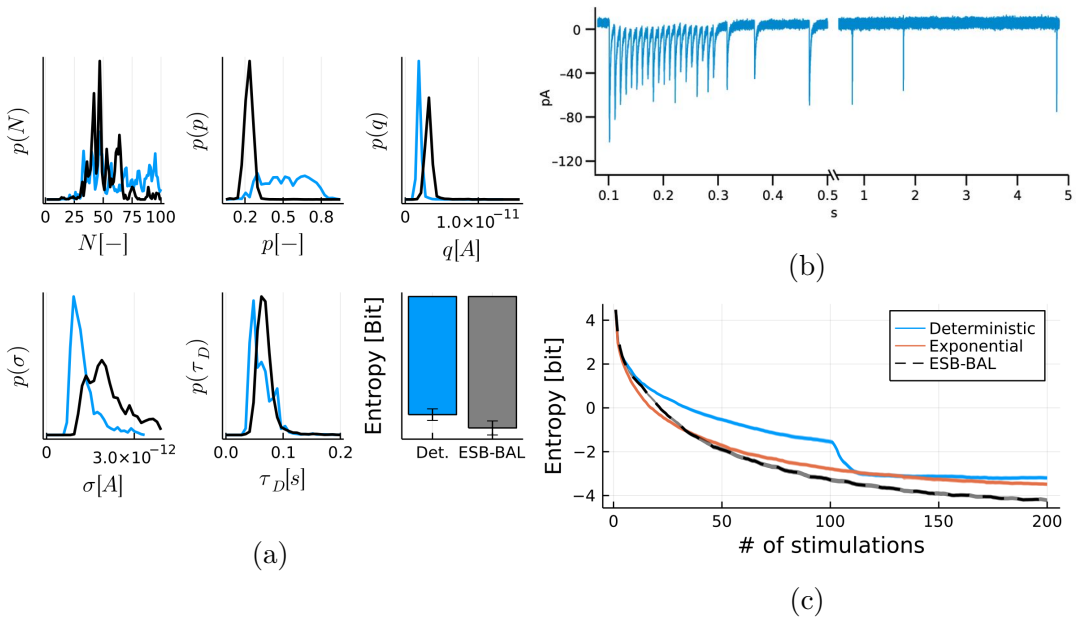


Figure 5.5: **Third setting: batch optimization and application to neural recordings**. (a) Marginal posterior distributions obtained after 200 stimulations of the same synapse using either the deterministic protocol (blue) or ESB-BAL (black). Posteriors were computed using the Metropolis-Hastings algorithm (50.000 samples). The bar plot indicates the mean differential entropy after 200 observations for 5 recording sessions using the deterministic protocol and 3 recording sessions using ESB-BAL on different synapses. Error bars show the standard error of the mean. Data from cerebellar mossy fiber to granule cell synapses. (b) Example of a postsynaptic current trace recorded using a tetanic stimulation phase followed by 6 recovery spikes. (c) Simulated experiment with ground-truth parameters $N^* = 55$, $p^* = 0.174$, $q^* = 4.86$ pA, $\sigma^* = 1.67$ pA, and $\tau_D^* = 0.0828$s (i.e. the MAP values from recordings shown in (a)).

We validate our method by applying it to EPSC recordings from acute mouse cerebellar slices. Figure 5.5 (a) shows the marginal posterior distributions for parameters of a cerebellar mossy fiber to granule cell synapse, obtained using either ESB-BAL or a deterministic protocol (consisting of repetitions of the same train of 100 stimuli at 100Hz followed by 6 recovery spikes). In Figure 5.5 (c), using simulated experiments, we also compare ESB-BAL to an Exponential stimulation protocol (in which ISIs are drawn from an exponential distribution, whose time constant is optimized as in Section 5.7.1), which has been proposed to provide better estimates of synaptic parameters as it spans a broader frequency space [159]. Our results show that batch optimization via ESB-BAL clearly outperforms deterministic stimulation.

5.8 Discussion

When designing an experiment in physiology, or when training a model on data in machine learning, it is common to choose a priori a fixed set of inputs to the studied system. The use of such non-adaptive, non-optimized protocols often leads to a large variance of the estimated parameters, even when using a large number of trials or data points. Bayesian active learning is an efficient method for optimizing these inputs, but exact solutions are often intractable and not applicable to online experiments.

Here, we introduce ESB-BAL, a novel framework combining particle filtering, parallel computing, and mean-field theory. ESB-BAL is general and sufficiently efficient to be applied to a wide range of settings. We use it to infer the parameters of a model of synapse: for this specific example, computation time is a critical constraint, since the typical ISI is shorter than 1s, and because several future inputs need to be optimized together. Using synthetic data and neural recordings, we show that our method has the potential to significantly improve the precision and speed of model-based inferences.

We expect active learning to be particularly beneficial to neurophysiology experiments involving live cells or subjects. By reducing the number of samples required to obtain a certain result, or by improving the efficiency of information gain, we can reduce the cost of the experiment and the need for animal subjects. A possible negative impact would be that improving the relative efficiency of neurophysiology experiments may lead to a larger field of applications and therefore a larger demand for animal experiments, analogously to Jevons Paradox [194].

Our approach has some room for improvements. An evident drawback of using particle filtering is that it requires a very large number of particles to provide low variance estimates, as the approximation error only decreases with the square root of the number of particles. Moreover, future experimental work should focus on implementing ESB-BAL for different and more complicated models of a chemical synapse, including for instance short-term facilitation [20, 5, 134, 159] or vesicle content variability [168, 1]. Finally, future theoretical work should focus on obtaining results on the convergence of the estimators when using active learning. When observations are independent and identically distributed (i.i.d.), active learning will give an unbiased estimate of the parameters, whose variance will decrease with the number of observations [195]. Such theoretical results lack for systems with correlated outputs (such as the EPSCs in the studied synapse model), possibly leading to information saturation [196] or biased estimates.

Overall, we expect our proposed solution to pave the way towards better estimates of stochastic models in neuroscience, more efficient training in machine learning, and more systematic and automated experimental designs. Especially, applying it to synaptic characterization would allow to better understand the role of synaptic transmission in learning and memory [197], to study the role of synaptic stochasticity [31, 138], and lead to an improved understanding of spike timing-dependent plasticity [198], synaptic homeostasis [32, 33], plasticity [34, 35, 36], and connectivity [199].

5.9 Supplementary material

5.9.1 Particle Filtering for synaptic characterization

Algorithm 2: Particle filtering [186] for computing one step update of the posterior distribution of parameters

Input: $\{\theta^i_{t-1}\}_{1\leq i\leq M_{\text{out}}}$, $\{n^{i,j}_{t-1}, k^{i,j}_{t-1}\}_{1\leq j\leq M_{\text{in}}}$, x_t, ψ_t ;
for i *in* $1 \dots M_{\text{out}}$ **do**
  Jittering: update the outer particles $\theta^i_t = \kappa(\theta^i_{t-1})$;
  for j *in* $1 \dots M_{\text{in}}$ **do**
    Propagation: Draw $n^{i,j}_t \sim p(n^{i,j}_t | n^{i,j}_{t-1}, k^{i,j}_{t-1}, \theta^i_t, \psi_t)$ and $k^{i,j}_t \sim p(k^{i,j}_t | n^{i,j}_t, \theta^i_t)$;
    Likelihood: compute $\tilde{w}^{i,j}_t = p(x_t | n^{i,j}_t, k^{i,j}_t, \theta^i_t)$;
  end
  Normalization: $\tilde{w}^{i,j}_t \leftarrow \tilde{w}^{i,j}_t / \sum_j \tilde{w}^{i,j}_t$;
  Inner particles resampling: resample $\{n^{i,j}_t, k^{i,j}_t\}_{1\leq j\leq M_{\text{in}}}$ based on $\{\tilde{w}^{i,j}_t\}_{1\leq j\leq M_{\text{in}}}$;
end
Compute $w^i_t = \frac{1}{M_{\text{in}}} \sum_j \tilde{w}^{i,j}_t$;
Normalization: $w^i_t \leftarrow w^i_t / \sum_i w^i_t$;
Outer particles resampling: resample $\{\theta^i_t\}_{1\leq i\leq M_{\text{out}}}$ and $\{n^{i,j}_t, k^{i,j}_t\}_{1\leq j\leq M_{\text{in}}}$ based on $\{w^i_t\}_{1\leq i\leq M_{\text{out}}}$;
Output: $\{\theta^i_t\}_{1\leq i\leq M_{\text{out}}}$, $\{n^{i,j}_t, k^{i,j}_t\}_{1\leq j\leq M_{\text{in}}}$

Computing the posterior distribution of θ also implies to specify a prior $p(\theta)$ from which the initial particles $\{\theta^i_0\}_{1\leq i\leq M_{\text{out}}}$ will be drawn. For simplicity, we consider here uniform priors (as in [5, 134]), although the algorithm readily extends to different choices of prior.

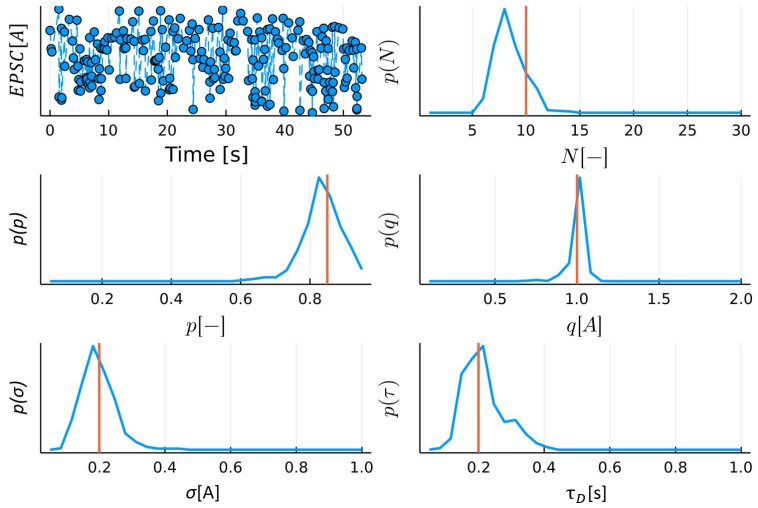


Figure 5.6: Examples of posteriors obtained using the filter (Algorithm 2). Upper left panel: train of synthetic EPSCs generated from the model described in Section 5.6. Other panels: posterior distributions of the parameters after 230 stimulations. Ground-truth values used to generate the EPSCs are displayed as red vertical lines.

5.9.2 Mean-field approximation of vesicle dynamics

Our synapse model, as defined by Eq. 5.7 to 5.8, is a Hidden Markov Model with observations x_t and hidden states n_t and k_t. The predictive distribution $p(x_{t+1}|x_{1:t}, \psi_{1:t+1}, \theta)$ used in Eq. 5.1 can be computed using the Baum-Welch algorithm: however, the algorithmic complexity of this forward-backward procedure, which scales with N^4, makes it impractical for closed-loop applications. Here, we suggest that computation can be massively simplified by using a mean-field approximation of vesicle dynamics: the analytical mean and variance of hidden and observed variables can be computed using recursive formulæ.

Let $r_t \in [0, 1]$ denote the average fraction of release-competent vesicles at the moment of spike t. Its values, given $\theta = [N, p, q, \sigma, \tau_D]$ and $\psi_{1:t}$, can be iteratively computed (see [20], Eq. (7)) from the equations of the Tsodyks-Markram model [17]: $r_t = 1 - (1 - (1 - p)r_{t-1}) \exp\left(-\frac{\psi_t}{\tau_D}\right)$ with $r_1 = 1$. It follows that the expected value of the EPSC after spike t is

$$\mathbb{E}(X_t|\psi_{1:t}, \theta) = r_t N p q \tag{5.10}$$

Similarly, the variance of the number of available vesicles $\mathrm{Var}(n_t)$ can be computed using the law of total variance: $\mathrm{Var}(n_t) = \mathbb{E}(\mathrm{Var}(n_t|n_{t-1}, k_{t-1})) + \mathrm{Var}(\mathbb{E}(n_t|n_{t-1}, k_{t-1}))$. Since $n_t = n_{t-1} - k_{t-1} + v_t$ with $v_t \sim \mathrm{Bin}(N - n_{t-1} + k_{t-1}, I_t)$, it follows that $\mathrm{Var}(n_t) = I_t(1 - I_t)N(1 - r_{t-1} + p r_{t-1}) + (1 - I_t)^2 \mathrm{Var}(n_{t-1} - k_{t-1})$.

Finally, by noting that $(n_t - k_t)|n_t \sim \mathrm{Bin}(n_t, 1 - p)$ and using again the law of total variance to compute $\mathrm{Var}(n_{t-1} - k_{t-1}) = \mathbb{E}(\mathrm{Var}(n_{t-1} - k_{t-1}|n_{t-1})) + \mathrm{Var}(\mathbb{E}(n_{t-1} - k_{t-1}|n_{t-1}))$, we obtain

$$\mathrm{Var}(X_t|\psi_{1:t}, \theta) = \sigma^2 + q^2(N r_t p(1 - p) + \mathrm{Var}(n_t)p^2) \tag{5.11}$$

5.9.3 First setting: reducing the uncertainty of estimates for a given number of observations

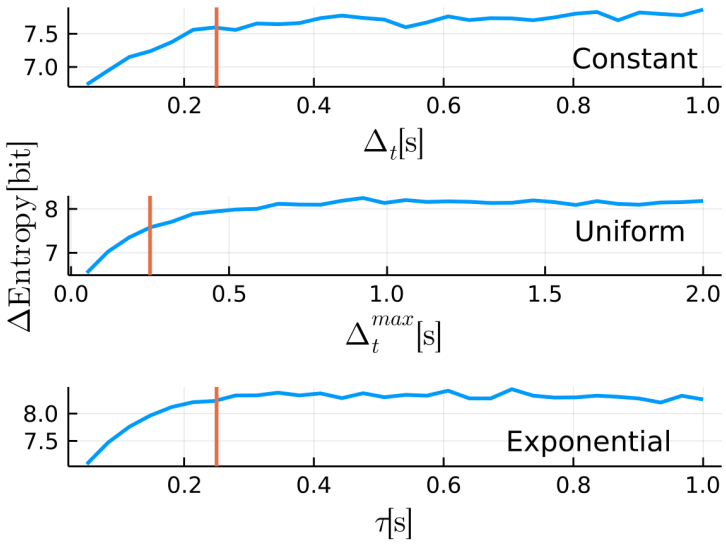
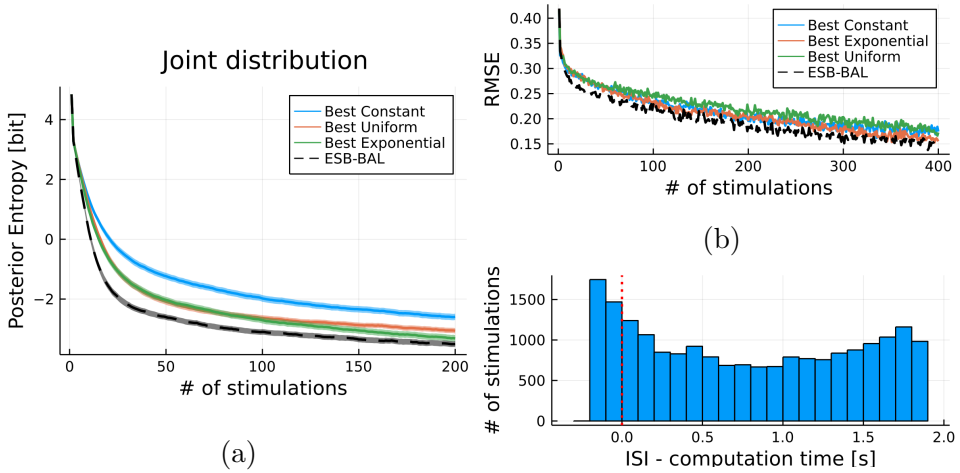
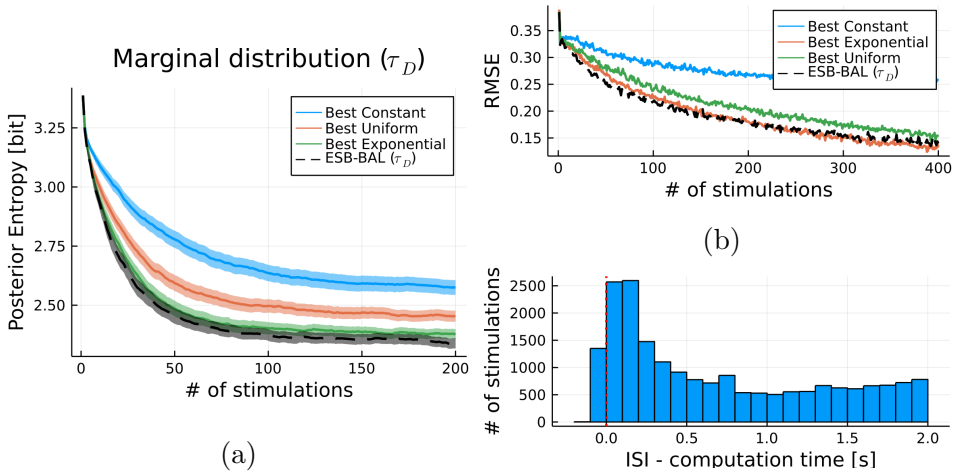


Figure 5.7: Average final entropy decrease (i.e. information gain) after 200 observations using the *Constant* (top), *Uniform* (middle), or *Exponential* (bottom) protocol, for different values of their hyperparameters. Ground truth parameters used are $N^* = 7$, $p^* = 0.6$, $q^* = 1$ A, $\sigma^* = 0.2$ A, and $\tau_D^* = 0.25$s [5]. Vertical red lines indicate the ground truth value $\tau_D^* = 0.25$s used for simulations. Optimal values for Δ_t, $\Delta_t^{\max}$, and τ are used in Figure 5.3.



(a) (b) (c)

Figure 5.8: Same setting as in Figure 5.3 but for ground truth parameters $N^* = 10$, $p^* = 0.85$, $q^* = 1$ A, $\sigma^* = 0.2$ A, and $\tau_D^* = 0.2$s.



(a) (b) (c)

Figure 5.9: Same setting as in Figure 5.3 but when optimizing solely for the marginal posterior distribution of τ_D.

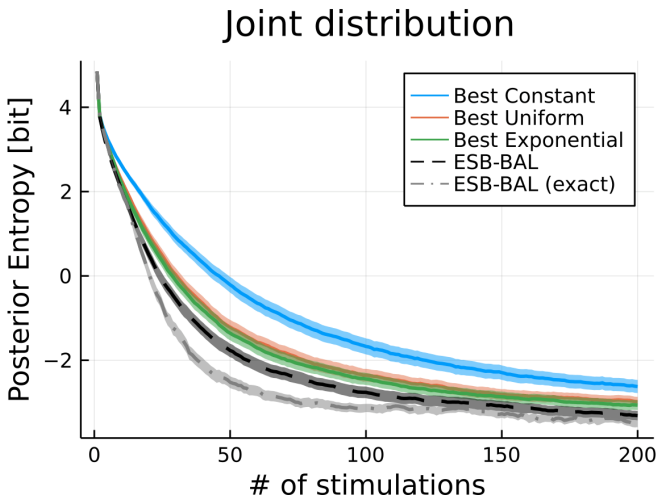


Figure 5.10: Same setting as in Figure 5.3, but with exact active learning (gray dashed line), in which Eq. 5.6 is computed exactly using MC samples.

5.9.4 Third setting: batch optimization and application to neural recordings

Algorithm 3: Computation of the optimal next batch of ISIs

set $\hat{\theta}_t = \arg\max_\theta p(\theta|h_t)$ (MAP values from the current posterior estimation);
Input: $\mathcal{S}_{t+1:t+n}$ (set of candidates $\psi_{t+1:t+n}$);
for $\psi_{t+1:t+n}$ *in* $\mathcal{S}_{t+1:t+n}$ **do**
  Compute $H_{h_t}^{\psi_{t+1:t+n}}(\Theta|X_{t+1:t+n})$ using Algorithm 2;
end
$\psi^*_{t+1:t+n} = \arg\min_{\psi_{t+1:t+n}\in\mathcal{S}_{t+1:t+n}} H_{h_t}^{\psi_{t+1:t+n}}(\Theta|X_{t+1:t+n})$

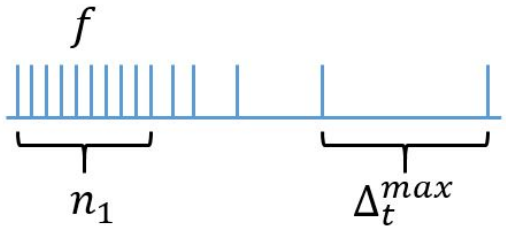


Figure 5.11: Schematic of how elements in $\mathcal{S}_{t+1:t+n}$ in Algorithm 3 are defined. Assuming that each element in $\mathcal{S}_{t+1:t+n}$ contains n ISIs, they are chosen so as to span 3 parameters: the number $n_1 < n$ of spikes in the tetanic stimulation phase, the frequency f of spikes in the tetanic stimulation phase, and the duration of the final recovery ISI Δ_t^{max}. The remaining $n_1 - 2$ spikes are then distributed geometrically between the end of the tetanic phase and the enultimate spike.

5.10 Acknowledgments and Disclosure of Funding

The work presented in this paper was supported by the Swiss National Science Foundation under grant number 31003A 175644 entitled "Bayesian Synapse".

Figures were created using BioRender. Calculations were performed on UBELIX (`http://www.id.unibe.ch/hpc`), the HPC cluster at the University of Bern. The CUDA.jl package [187, 188] is licensed under the MIT "Expat" License (`https://github.com/JuliaGPU/CUDA.jl/blob/master/LICENSE.md`). We thank Ehsan Abedi, Jakob Jordan, and Anna Kutschireiter for the fruitful discussions.

`Julia` files are available in the following package:

`https://github.com/Theoretical-Neuroscience-Group/BinomialSynapses.jl`

Chapter 6

Optimal experiment design as identifiability domain enhancement

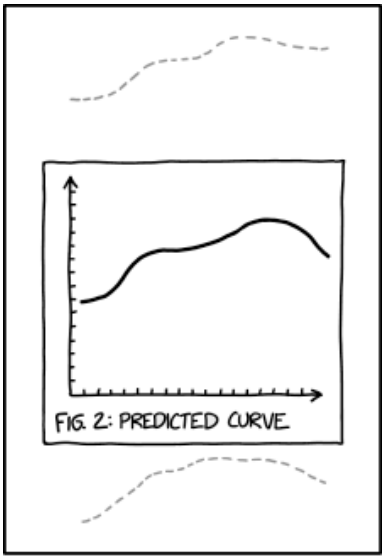


Figure 6.1: On thoughtful model selection [1].

6.1 Introduction

In the previous Chapter (Chapter 5), we introduced a method called Efficient Sampling-Based Bayesian Active Learning (ESB-BAL), which can be used for optimizing sequential experiments. Using particle filtering, ESB-BAL selects the next experimental design to maximize the statistical mutual information between the output of the experiment and the constants of the studied system. To validate it, we applied ESB-BAL

[1] `https://xkcd.com/2311/`, use with authorization.

to the problem of estimating the constants of a chemical synapse from its postsynaptic currents evoked by presynaptic stimulations. After each new observation, the optimal next stimulation time can be computed using ESB-BAL. Using synthetic data and synaptic whole-cell patch-clamp recordings in cerebellar brain slices, we show that our method is efficient enough to be used in real-time biological experiments and can significantly reduce the uncertainty of inferred parameters.

However, ESB-BAL aims at answering a specific question: given a certain model $\mathcal{M}$, how to select the experimental design Ψ that will minimize the entropy of the posterior distribution of the parameters $p(\theta|\mathcal{D}, \mathcal{M})$? The main assumption is that $\mathcal{M}$ is known: ESB-BAL thus solves **Optimal experiment design for parameters inference**, but not **Optimal experiment design for model selection** (see Section 1.8). Indeed, in practice a model selection step is often required prior to parameter inference to discriminate among a set of candidate models and to find which $\mathcal{M}$ to fit on the data. OED can also be used for optimal model selection: how to select the experimental design Ψ that will minimize the entropy of the posterior distribution over the models $p(\mathcal{M}|\mathcal{D})$?

In the following sections, we will describe classical methods for performing OED for model selection, and propose an alternative approach, which aim is to maximize the identifiability (in the sense of Chapter 4) of an alternative model $\mathcal{M}_1$ compared to a null model $\mathcal{M}_0$. Classical methods for optimal model selection, which are agnostic to the complexity of the competing models, suffer from several pitfalls, including large algorithmic complexity, difficulty to design the set of competing models, and sensitivity to local optima. In contrast, our proposed approach, which we call Identifiability Domain Enhancement (IDE), does not aim at minimizing the entropy of the distribution over the models. Rather, its aim is to maximize the informativeness of an experiment and the identifiability of a studied model: this allows experimentalists to draw more robust conclusions from their experiments, by increasing the discriminability between two competing candidate models and improving the accuracy of their estimated parameters.

6.2 Classical OED for model selection

As mentioned previously, two different setups for Optimal Experiment Design need to be distinguished. Firstly, **OED for parameter inference**, in which the cost function (that needs to be minimized by picking the optimal design Ψ^*_{param}) is usually defined as the entropy of the posterior distribution of the parameters for a given model $\mathcal{M}$:

$$\Psi^*_{\text{param}} = \arg\min_{\Psi} H(\theta|\mathcal{D}, \mathcal{M}, \Psi) \tag{6.1}$$

Secondly, **OED can also be used for model selection**. In this case, the cost function is classically defined as the entropy of the distribution over the models [200, 201, 202]:

$$\Psi^*_{\text{model}} = \arg\min_{\Psi} H(\mathcal{M}|\mathcal{D}, \Psi) \tag{6.2}$$

Other formulations of OED for model selection have also been proposed. In [82], the authors propose to use the Jensen-Shannon divergence between the densities of the

M competing models to optimize their discriminability:

$$\Psi^*_{\text{model}} = \arg\max_{\Psi} \sum_{i=1}^{M} p(\mathcal{M}_i) D_{KL}\left(p(\mathcal{D}|\mathcal{M}_i,\Psi)||\sum_{i=1}^{M} p(\mathcal{M}_i)p(\mathcal{D}|\mathcal{M}_i,\Psi)\right) \tag{6.3}$$

In [83], the authors used a one-sided form of the Kullback-Leibler divergence, hence computing M designs $(\Psi^*_i)_{1\leq i\leq M}$ optimizing the discriminability of model $\mathcal{M}_i$ against the $M-1$ competing models:

$$\Psi^*_i = \arg\max_{\Psi} \int d\mathcal{D} p(\mathcal{D}|\mathcal{M}_i,\Psi) \log\left(\frac{(1-p(\mathcal{M}_i))p(\mathcal{D}|\mathcal{M}_i,\Psi)}{\sum_{j\neq i} p(\mathcal{M}_j)p(\mathcal{D}|\mathcal{M}_j,\Psi)}\right) \tag{6.4}$$

Other utility functions, such as the Zero–One utility, Ds-optimality [200], or estimating the expected loss via a supervised classification procedure [201] have also been proposed.

Interestingly, no formal link between OED for model selection and OED for parameter inference exists. Since the cost functions in Eq. 6.1 and 6.2 are different, Ψ^*_{param} and Ψ^*_{model} are not necessarily equal, i.e. a design that will be optimal for model selection will not be optimal for parameter inference. One can remark that the joint entropy of the model random variable $\mathcal{M}$ and of the parameters random variable θ can be expressed as

$$H(\mathcal{M},\theta|\mathcal{D},\Psi) = H(\mathcal{M}|\mathcal{D},\Psi) + H(\theta|\mathcal{M},\mathcal{D},\Psi)$$

where the first term on the right-hand side corresponds to the entropy to be minimized in model selection, and the second term corresponds to the entropy to be minimized in parameter inference. It is thus possible to optimize either this total entropy, or a weighted sum of the terms on the right-hand side [203, 204, 202]. However, both approaches are often considered to be separated problems [82, 83, 201, 200].

These classical formulations of OED for model selection have several important drawbacks:

- Computing the cost function for selection of the optimal design Ψ^*_{model} can be computationally challenging and lead to important run times, as illustrated in [200];
- In OED for parameter inference, specifying a range or a prior distribution for the candidate parameters θ is usually fairly simple. However, defining a discrete family of candidate models is not always straightforward[2], as it is not always possible to define a similarity or distance metric between different models or to define a space of all possible models;
- If the state of possible experimental parameters is ill-defined, OED for model selection can lead to an experimental design which successfully minimizes the cost function (e.g. the entropy of $p(\mathcal{M}|\mathcal{D},\Psi)$) but selects the wrong model (as illustrated in Figure 6.2).

[2]See for instance `https://stats.stackexchange.com/a/494919/271601`

We here consider a different approach: instead of minimizing the entropy (or a related metric) of the posterior distribution over a set of candidate models, our aim is to maximize the identifiability (in the sense of Chapter 4) of an alternative model $\mathcal{M}_1$ compared to a null hypothesis $\mathcal{M}_0$.

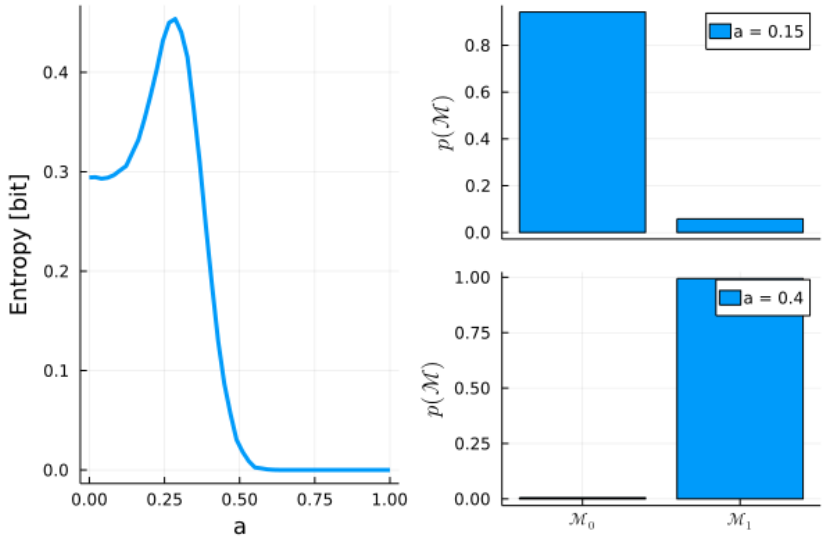


Figure 6.2: **Local optimum in classical OED for model selection**. Data are generated from a ground truth model $\mathcal{M}_1$ where the (scalar) input x and output y are linked as $y = \alpha x^2 + \beta x + \epsilon$, with $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Two competing models are compared: a constant model $\mathcal{M}_0$ according to which $y = \gamma$ (one free parameter) and the ground truth model $\mathcal{M}_1$ (two free parameters). Surrogate data are generated from $\mathcal{M}_1$ using a vector of values for x containing 50 equally separated values between $-a$ and a. The goodness of fit of $\mathcal{M}_0$ and $\mathcal{M}_1$ for these data is compared for different values of a. The left panel shows the average entropy of the distribution $p(\mathcal{M}|\mathcal{D})$ for 5000 independent realizations of x with $\sigma = 0.1$. For $a \approx 0$, both models have approximately the same likelihood (since they have similar outputs for $x \approx 0$), but $\mathcal{M}_1$ is penalized for its complexity (upper right panel). As a increases, so does $p(\mathcal{M}_1|\mathcal{D})$, and the ground truth model is correctly selected (lower right panel). If the parameter space is poorly chosen (e.g. if a is constrained within $[0; 0.25]$), minimizing the entropy of $p(\mathcal{M}|\mathcal{D})$ would yield a locally optimal experimental design ($a \approx 0$) which would favor the wrong model.

6.3 Identifiability Domain Enhancement

Statistical hypothesis testing, i.e. comparing the null hypothesis $\mathcal{H}_0$ and the alternative hypothesis $\mathcal{H}_1$ based on observations $\mathcal{D}$, can be reformulated as a model selection problem, i.e. computing whether the null model $\mathcal{M}_0$ or the alternative model $\mathcal{M}_1$ is a better fit to the data $\mathcal{D}$. $\mathcal{M}_0$ is often assumed to be a submodel of $\mathcal{M}_1$. This approach allows to draw conclusions on the nature of the studied synapse, e.g.:

- If a binomial model $\mathcal{M}_1$ provides a better fit to the data than a Gaussian model $\mathcal{M}_0$, this indicates that synaptic transmission is likely to be quantised;
- If a model with STP $\mathcal{M}_1$ provides a better fit to the data than a static model $\mathcal{M}_0$, this indicates that synaptic transmission is likely to be dynamic.

In this context, Optimal Experiment Design becomes critical to avoid identifiabil-

ity issues. Indeed, some features of the studied synapse (e.g. the above-mentioned quantal release or short-term plasticity) may not be experimentally observable, depending on the experimental protocol being used. Large recording noise, or poorly chosen stimulation protocols, could make interesting features of the studied system non-observable and may not allow to draw conclusions on the nature of the synapse. When a simpler model provides a better fit (in the sense of model selection) than a more complicated alternative release model, it is impossible to conclude whether it is due to the intrinsic nature of the synapse or to non-observability.

An intuitive solution to prevent this pitfall is to design the experimental protocol Ψ in order to maximize the identifiability (in the sense of Chapter 4) of the alternative hypothesis $\mathcal{M}_1$ compared to submodel $\mathcal{M}_0$. In Section 4.3.6, we proposed a definition for the practical identifiability of a statistical model: for a given parametrization θ and a given set of stimulations Ψ (i.e. the experimental design), the model $\mathcal{M}_1$ is said to be practically identifiable with respect to submodel $\mathcal{M}_0$ if the following condition is met:

$$\langle \log p(\mathcal{D}|\mathcal{M}_1, \Psi) \rangle_{p(\mathcal{D}|\theta,\mathcal{M}_1,\Psi)} \geq \langle \log p(\mathcal{D}|\mathcal{M}_0, \Psi) \rangle_{p(\mathcal{D}|\theta,\mathcal{M}_1,\Psi)} \tag{6.5}$$

The identifiability domain of $\mathcal{M}_1$ corresponds to all the settings (parameters and protocols) for which, on average, data generated from the ground truth $\mathcal{M}_1$ will be better explained by $\mathcal{M}_1$ than by its submodel. Inequality 6.5 depends on θ, i.e. represents the identifiability of $\mathcal{M}_1$ compared to $\mathcal{M}_0$ for a specific value θ of the random variable Θ. For instance, Figure 4.5 (panels A and B) shows the identifiability domain of a binomial model compared to a unimodal model for different values of p and σ. Similarly, Figure 4.6 shows the identifiability domain of a model with STD compared to a static model for different values of p and τ_D.

By averaging 6.5 over $p(\theta|\mathcal{M}_1)$ (i.e. the prior distribution of θ)[3] we define a global identifiability level $f(\Psi, \mathcal{M}_1, \mathcal{M}_0)$ for $\mathcal{M}_1$ compared to $\mathcal{M}_0$:

$$f(\Psi, \mathcal{M}_1, \mathcal{M}_0) = \int d\mathcal{D} p(\mathcal{D}|\mathcal{M}_1, \Psi) \log p(\mathcal{D}|\mathcal{M}_1, \Psi) - \int d\mathcal{D} p(\mathcal{D}|\mathcal{M}_1, \Psi) \log p(\mathcal{D}|\mathcal{M}_0, \Psi) \tag{6.6}$$

which corresponds to the Kullback-Leibler divergence between the competing model evidences $p(\mathcal{D}|\mathcal{M}_1, \Psi)$ and $p(\mathcal{D}|\mathcal{M}_0, \Psi)$, and is thus always positive:

$$f(\Psi, \mathcal{M}_1, \mathcal{M}_0) = D_{KL}(p(\mathcal{D}|\mathcal{M}_1, \Psi)||p(\mathcal{D}|\mathcal{M}_0, \Psi)) \geq 0 \tag{6.7}$$

Intuitively, picking the experimental design Ψ which will maximize the value of $f(\Psi, \mathcal{M}_1, \mathcal{M}_0)$ can be thought of as maximizing the identifiability domain (as defined in Chapter 4) of $\mathcal{M}_1$, and hence reducing identifiability issues when performing model selection. In the following sections, we apply this Identifiability Domain Enhancement approach to different synaptic models, and empirically assess its effect on parameter inference.

[3] Note that the prior distribution over the parameters does not depend on the experimental design Ψ.

6.4 Application to quantal release

Our first example of application is to compare the identifiability of the classical binomial model with parameters $[N, p, q, \sigma]$ (denoted here as $\mathcal{M}_1$) and of a simple Gaussian distribution with parameters $[\mu, \sigma]$ (denoted here as $\mathcal{M}_0$), as in Figure 4.5. In this example, none of the studied models account for short-term plasticity: inter-spike intervals are thus irrelevant for these models, and the only experimental variable at hand in the protocol Ψ is the number of observations T.

It can be seen from the identifiability domain plots that, for a given number of observations T, $\mathcal{M}_1$ is only identifiable if the recording noise is sufficiently low with respect to the quantal amplitude q for subpeaks in the distributions of EPSCs to be observable. Similarly, for a given recording noise, the identifiability of $\mathcal{M}_1$ will increase with T. This can be understood intuitively: as the number of independent observations increases, so will the possibility to accurately infer the ground truth model and its parameters. The best experimental design Ψ^* that will maximize the identifiability level of $\mathcal{M}_1$ (Eq. 6.7) is thus the one in which T is as high as possible.

Besides, increasing T is also optimal for parameter inference, i.e. for minimizing the entropy of the posterior distribution of the parameters. As explained in [195], if $\mathcal{M}$ is a model which outputs are i.i.d., the posterior distribution $p(\theta|\mathcal{D}, \mathcal{M})$ will be asymptotically normal, with a mean converging to the ground truth value of θ (i.e. the solution will be unbiased) and a variance decreasing as $\frac{1}{T}$ as more data points are acquired.

Although intuitive, this result can also be derived from a Bayesian Active Learning argument. The classical chain rule for entropy yields

$$H(\theta, \mathcal{D}|\Psi) = H(\mathcal{D}|\Psi) + H(\theta|\mathcal{D}, \Psi) \tag{6.8}$$

If observations in $\mathcal{D} = \{e_i\}_{1 \leq i \leq T}$ are i.i.d., that is to say if all e_i are independent realizations of the same random variable e, then

$$H(\mathcal{D}|\Psi) = T \cdot H(e) \tag{6.9}$$

i.e. the entropy $H(\mathcal{D}|\Psi)$ increases linearly with the number of observations with a factor $H(e)$. Similarly,

$$H(\mathcal{D}, \theta|\Psi) = H(\theta) + T \cdot H(e|\theta) \tag{6.10}$$

Since $H(X|Y) \leq H(X)$ for any random variables X and Y, Equations 6.9 and 6.10 yield

$$\mathbb{E}_\theta(H(e|\theta)) \leq H(e) \tag{6.11}$$

Combining Equations 6.8 and 6.11 shows that increasing T minimizes the last term in the right-hand side of Eq. 6.8, i.e. minimizes the expected entropy of the posterior distribution of θ. In conclusion, for model where observations are i.i.d. and where the number of observations T is the only experimental parameter in Ψ, increasing T is optimal both for the identifiability level of $\mathcal{M}_1$ and for parameter inference.

6.5 Application to Short-Term Depression

Our second example of application is to compare the identifiability of the binomial model with short-term depression (i.e. with parameters $[N, p, q, \sigma, \tau_D]$), denoted here as $\mathcal{M}_2$, and of the binomial model without STD $\mathcal{M}_1$ (i.e. with parameters $[N, p, q, \sigma]$). An interesting parameter that can be tuned by the experimentalist is the calcium concentration $[Ca^{2+}]$ in the experimental solution, which will affect the synapse's release probability p. Indeed, a higher calcium concentration will increase the vesicles' probability to fuse with the plasma membrane, while a low $[Ca^{2+}]$ increases the probability of synaptic failures [9, 205]. In this section, we will discuss how $[Ca^{2+}]$ can be tuned for optimizing the experimental design.

Figure 4.6 shows the identifiability domain of $\mathcal{M}_2$ as a function of p and τ_D for a given stimulation protocol. One can remark that STD is more easily observed for higher values of p. This can be intuitively understood: depression can only be observed when the release probability is sufficiently high to induce vesicle pool depletion. For $p \approx 1$, stochasticity at the level of vesicles release disappears (since $\forall i \; k_i = n_i$), thus allowing to more accurately estimate the binomial parameters and the refilling dynamics time constant.

Besides, increasing the external calcium concentration, and hence the release probability p, is also beneficial for parameter inference, i.e. for minimizing the entropy of $p(\theta|\mathcal{D}, \mathcal{M}_2)$. Figure 6.3 shows the average posterior entropy of the parameters (averaged over 100 independent repetitions and computed using the NPF described in Chapter 5) for different values of the ground-truth parameter p. As expected, increased $[Ca^{2+}]$ leads to more accurate parameter inference for $\mathcal{M}_2$. The reduced posterior entropy for higher $[Ca^{2+}]$ can also be visually assessed when using the Metropolis-Hastings algorithm to compute the marginal posterior distributions of N, p, q, σ, and τ_D for different ground-truth values of p (Figure 6.4).

Tuning the value of the experimental parameter $[Ca^{2+}]$ thus allows to optimize the identifiability of $\mathcal{M}_2$ and the inference of its parameters.

6.6 Application to variable quantal amplitudes

An important assumption of the classical binomial model $\mathcal{M}_1$ is that the quantal amplitudes q are homogeneous and identical for each release site. The probability distribution of EPSCs is thus defined by Eq. 1.1: it is a mixture of Gaussian (Figure 1.9, A) where each component:

- has its mean located to a multiple $k_i q$ with k_i in $\{0, ..., N\}$;
- has a width corresponding to the recording noise σ;
- has its height determined by the latent variable probability $p(k_i) = \text{Bin}(k_i; N, p)$.

However, mEPSC amplitudes have been observed not to be exactly similar and stereotypical, but rather to follow a unimodal distribution [1]. The effect of one release event q can thus be thought of as a random variable instead of a constant. Following [168] and [1], we can assume that q follows a gamma distribution with shape parameter α and scale parameter[4] β, i.e. $q \sim \Gamma(\alpha, \beta)$. In fine, we can define an alternative binomial model $\mathcal{M}_1'$ where each EPSC e_i is defined as $e_i = \gamma_i + \epsilon_i$, with:

[4] An alternative parametrization of the gamma distribution is also possible, with a shape parameter α and an inverse scale (or rate) parameter $1/\beta$.

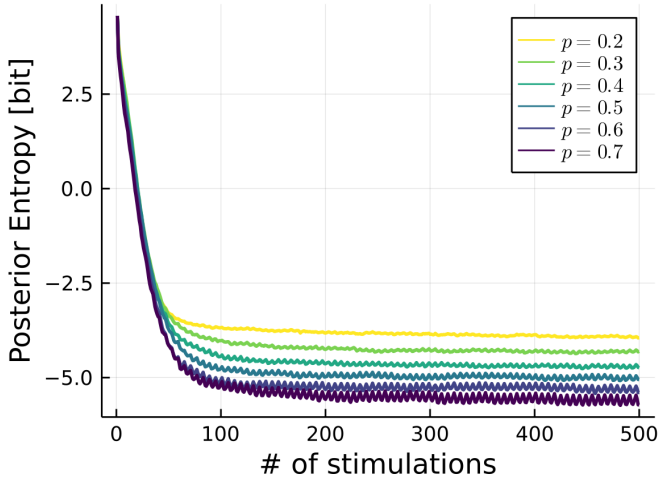


Figure 6.3: **Increased $[Ca^{2+}]$ leads to more accurate parameter inference for $\mathcal{M}_2$ (NPF).** The entropy of the posterior distribution $p(\theta|\mathcal{D}, \mathcal{M}_2)$ is lower for higher values of $[Ca^{2+}]$ (i.e. higher values of p). Surrogate data were generated from $\mathcal{M}_2$ with parameters $N = 5$, $q = 1$, $\sigma = 0.2$, $\tau_D = 0.25$, and a stimulation protocol consisting in repetitions of the same spike train (4 spikes with an inter-spike interval $ISI = 0.05s$ followed by a recovery spike $0.5s$ later), i.e. the same setting as in Figure 4.6 (B). Solid lines show, for different values of the ground-truth parameter p, the entropy of $p(\theta|\mathcal{D}, \mathcal{M}_2)$ averaged over 100 independent repetitions and computed using the NPF (see Chapter 5).

$$k_i \sim \text{Bin}(N, p) \tag{6.12}$$

$$\gamma_i|k_i \sim \Gamma(k_i\alpha, \beta) \tag{6.13}$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2) \tag{6.14}$$

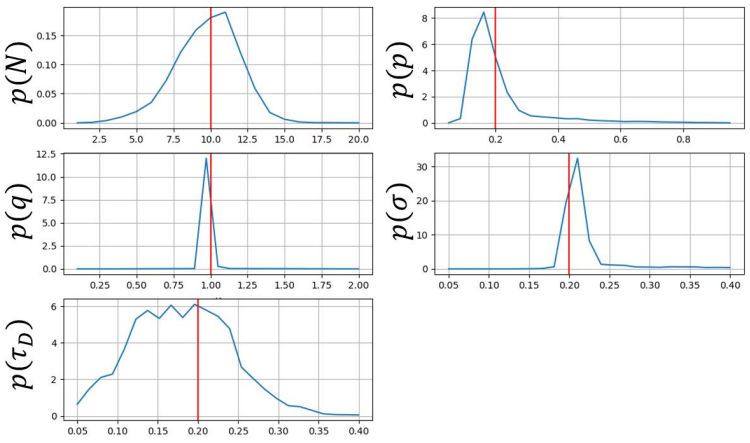
(i.e. $e_i|\gamma_i \sim \mathcal{N}(\gamma_i, \sigma^2)$). The probability of a given observation e_i can thus be computed as

$$p(e_i) = \sum_{k_i} \int d\gamma_i p(e_i|\gamma_i)p(\gamma_i|k_i)p(k_i) \tag{6.15}$$

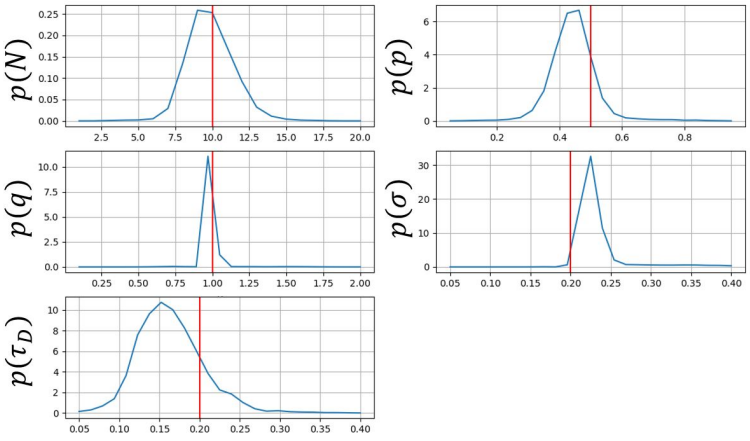
In [168], the authors state that for $k_i \geq 1$ the variability due to the recording noise will be negligible compared to the variability due to the quantal components, and can thus be ignored. They simplify the computation of Eq. 6.15 into

$$p(e_i) = \text{Bin}(0; N, p)\mathcal{N}(e_i; 0, \sigma^2) + \sum_{k_i=1}^{N} \text{Bin}(k_i; N, p)\Gamma(e_i; k_i\alpha, \beta) \tag{6.16}$$

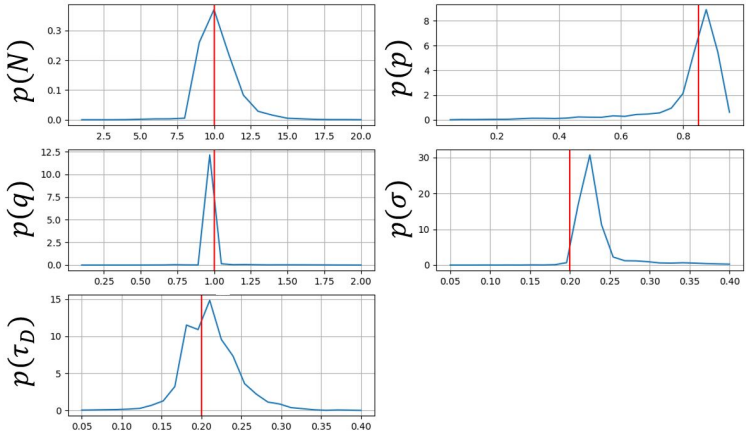
This approximation is used in Figure 6.5, hence explaining the small discrepancy between the theoretical distribution (orange plot) and the observed empirical distribution (histogram of samples).



(a)

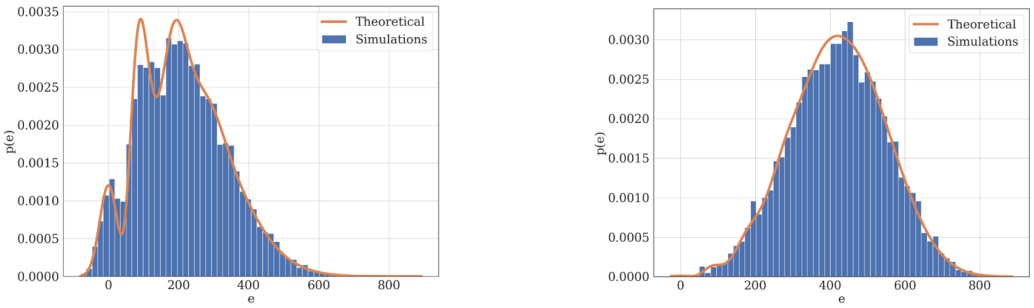


(b)



(c)

Figure 6.4: **Increased $[Ca^{2+}]$ leads to more accurate parameter inference for $\mathcal{M}_2$ (MH)**. Three independent sets of 100 EPSCs were generated from $\mathcal{M}_2$ for different ground-truth values of p ($p = 0.2$ in (a), $p = 0.5$ in (b), $p = 0.85$ in (c)). The marginal posterior distributions of their parameters was then computed using the Metropolis-Hastings algorithm. A higher variability (especially for the distributions of N and τ_D) can be visually verified for lower values of p. Surrogate data were generated from $\mathcal{M}_2$ with parameters $N = 5$, $q = 1$, $\sigma = 0.2$, $\tau_D = 0.25$, and a stimulation protocol consisting in repetitions of the same spike train (4 spikes with an inter-spike interval $ISI = 0.05s$ followed by a recovery spike $0.5s$ later), i.e. the same setting as in Figure 4.6 (B).



(a) (b)

Figure 6.5: **Theoretical and empirical distributions of EPSCs under $\mathcal{M}_1'$.** (a) Blue: histogram of 10000 surrogate EPSCs generated from $\mathcal{M}_1'$ (Equations 6.12 to 6.14) with ground-truth parameters $N = 5$, $p = 0.35$, $\alpha = 11.11$, $\beta = 0.11$, and $\sigma = 25$ (i.e. same values as in [168]). Orange: theoretical probability distribution of EPSCs (Eq. 6.16). The distribution is a mixture of Gamma components, which width increases linearly with the number of released vesicles k_i. (b) Same setting, but for $p = 0.7$. Components have fused into a unimodal distribution, impeding the possibility to identify quantal transmission.

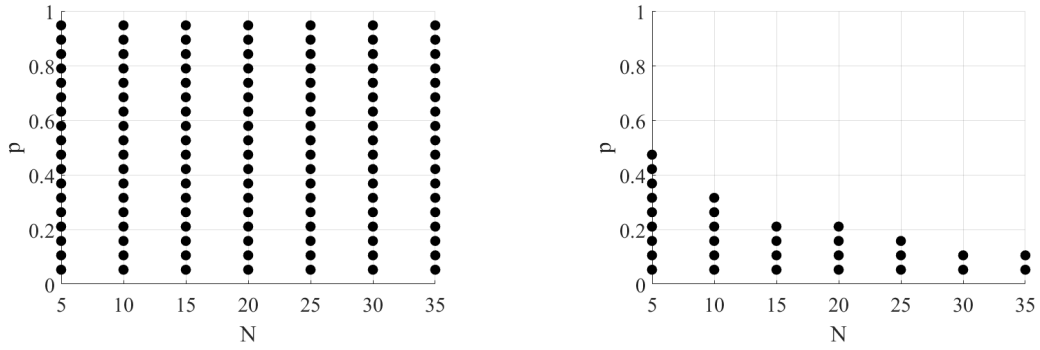
Under $\mathcal{M}_1'$, the distribution of EPSCs is a mixture of Gamma components, which width increases linearly with k_i (Figure 6.5 (A)). If p is large, these components will fuse into a unimodal distribution, impeding the possibility to identify quantal transmission and to compute the parameters (Figure 6.5 (B)).

If quantal amplitudes are indeed variable (which is corroborated by experimental observations [168, 1]), then the quantal release of neurotransmitters can only be observed for low values of the release probability p. As a matter of fact, the seminal paper by J. Del Castillo and B. Katz [9], in which the quantal nature of neurotransmitters release was first suggested from the multimodal distribution of postsynaptic potentials, used a lowered external calcium concentration. Similarly, in [205], the multimodal distribution of postsynaptic currents is only observed at low $[Ca^{2+}]$ (see Figures 1-C and 5 in [205]). This intuition can be formalized by studying the identifiability domain of $\mathcal{M}_1'$ compared to $\mathcal{M}_0$ and plotted as a function of p and N (Figure 6.6): panel B shows how the identifiability of $\mathcal{M}_1'$ decreases at larger p.

6.7 Conclusion

To conclude, we have proposed here a different approach to design an experiment when performing model selection. Contrary to classical methods for OED for model selection, which aim is to minimize the entropy of the distribution $p(\mathcal{M}|\mathcal{D}, \Psi)$, our method, which is called Identifiability Domain Enhancement, uses the notion of identifiability domain presented in Chapter 4. In this approach, the experimental design Ψ is chosen so as to maximize a global identifiability level for the tested model which is compared to a null model.

Generally, the goal of an experiment is to propose a new, alternative, hypothesis $\mathcal{H}_1$, and to compare it to the previous hypothesis $\mathcal{H}_0$. This process can be formulated as a model selection problem: the goal of an experiment is to assess whether the proposed model $\mathcal{M}_1$ provides a better fit to observations $\mathcal{D}$ than the null model $\mathcal{M}_0$.



(a) (b)

Figure 6.6: **Identifiability domains of $\mathcal{M}_1$ and $\mathcal{M}_1'$.** (a) Identifiability domain of $\mathcal{M}_1$ (compared to $\mathcal{M}_0$) plotted as a function of p and N. Results obtained for $T = 500$, $q = 1$, $\sigma = 0.2$. Same computation method as in Section 4.4.1. (b) Identifiability domain of $\mathcal{M}_1'$ (compared to $\mathcal{M}_0$) plotted as a function of p and N. Results obtained for $T = 500$, $\alpha = 11.11$, $\beta = 0.11$, $\sigma = 25$ (i.e. same values as in [168]).

But doing so requires observations to be sufficiently informative about $\mathcal{M}_1$. Our proposed approach to experimental design aims at making sure the features of $\mathcal{M}_1$ are not "hidden".

Besides, we have empirically shown for different examples (Section 6.4 for quantal release, Section 6.5 for STD) that improving this global identifiability metric also improves parameter inference, i.e. reduces the entropy of the posterior distribution $p(\theta|\mathcal{D}, \mathcal{M})$. More generally, identifiability domains can be used as a tool for experimentalists to tune their experimental parameters (e.g. how to set the value for $[Ca^{2+}]$ as seen in Figure 6.6).

Chapter 7

A normative approach to synaptic proteins organization

In this chapter, we present preliminary results on modeling synaptic transmission based on subsynaptic protein clusters distribution, which were presented at the Champalimaud Research Symposium 2021[1]. In previous chapters, we focused on high-level models of chemical synapses and on evoked synaptic transmission. Here, we propose to study chemical synapses at a finer level. Using both microscopy observations and recordings of spontaneous postsynaptic currents, we derive models of synaptic protein organization and of postsynaptic receptors sensitivity to neurotransmitter.

An interesting biological question to which we ought to bring a theoretical answer is how to make sense of the observed organization of synaptic proteins. Synaptic transmission critically depends on the relative positions of presynaptic neurotransmitter release sites and of postsynaptic receptors. Structural alignments of synaptic proteins, called transsynaptic nanocolumns [206], had been previously observed using localization microscopy. More specifically, gSTED microscopy data indicates a ring arrangement of transsynaptic nanocolumns at the larval Drosophila neuromuscular junction.

In [207], the authors studied the size of neuronal structures. Using an information transmission argument, they derived what should the optimal axon diameter be, and compared this normative prediction to observed neuron sizes. Here, we use a similar approach: we propose a normative approach to this protein organization, and suggest that it might optimize a certain biological cost function (e.g. the mean current or SNR after vesicle release). As in [208], the goal is to use models of the NMJ to validate experimental hypotheses about synaptic transmission.

7.1 Introduction

There is evidence that synaptic proteins arrange in sub-synaptic clusters that may organize across the synaptic cleft. Superresolution microscopy allows to observe both presynaptic BRP clusters, and postsynaptic glutamate receptors (GluR) clusters. BRP are a presynaptic scaffold, and Unc13 clusters, which may be a proxy for the

[1] C. Gontier, W. Gauthey, S. Sydlik, M. Müller, and J.-P. Pfister. Modeling synaptic transmission based on subsynaptic glutamate receptors distribution. Champalimaud Research Symposium 2021, Lisbon, Portugal.

position of release sites [209], localize close to their C-termini. Data suggest that presynaptic nano-clusters and postsynaptic GluRs form structures that arrange in stereotypical ring patterns at the Drosophila neuromuscular junction (NMJ) [11]. However, this model of transsynaptic rings has not been tested against other models. Moreover, the relationship between this transsynaptic nanoarchitecture and synaptic function remains elusive. Here, we employed a theoretical approach to test whether pre- and postsynaptic nano-clusters indeed arrange in transsynaptic rings, and to relate this transsynaptic nano-organization to synaptic transmission. Moreover, we investigate transsynaptic nano-architecture in relation to synaptic transmission after genetic ablation of Unc13A, a gene encoding a protein suggested to be the biological correlate of presynaptic release sites. We address the following questions:

1. How are BRP and GluR clusters spatially organized ?
2. Can we define a simple model of GluR sensitivity to glutamate ?
3. Based on the organization of BRP and GluR clusters, and on the sensitivity of GluR clusters to glutamate, can we explain the distribution of observed spontaneous excitatory postsynaptic currents (mEPSCs) ?
4. How does the predicted optimal position of BRP clusters compare to observations ?

7.2 Description of the experimental apparatus

The experimental apparatus is described in Figure 7.1. Observations were obtained from the Drosophila neuromuscular junction. Details regarding the experimental model and subjects can be found in [11]. A single NMJ corresponds to one single axon with several and heterogeneous AZ on it, all linked to a muscle cell at isopotential. Two kinds of observations are used:

- **Microscopy observations** obtained using gSTED microscopy [210]. It allows to determine, for several active zones, the position of presynaptic BRP proteins (which are used as a proxy for the position of release sites) and of postsynaptic GluR clusters.
- **Electrophysiology observations** obtained using voltage-clamp. It allows to record mEPSCs due to the spontaneous release of vesicles at active zones. The two-electrode voltage-clamp methodology being used is described in [33].

The average number of AZ per NMJ has been observed to be 128 in WT subjects and 58 in Unc13A mutants. In Unc13A mutants, 40% of the total number of AZ are made up of enlarged AZ, the other ones being similar to WT active zones.

7.3 Model comparison supports ring-like structures that are disrupted in Unc13A mutants

Our first step is to derive a quantitative description of proteins distribution, and hence to propose a generative model for the joint localization of presynaptic BRP proteins and postsynaptic GluR proteins. The position of observed BRP and GluR clusters can be understood as a multitype point-process, where each point is assigned a position (x_1, x_2) on the 2-dimensional AZ as well as a type (either BRP or $GluR$). We propose different models for this point process, each having different intensity

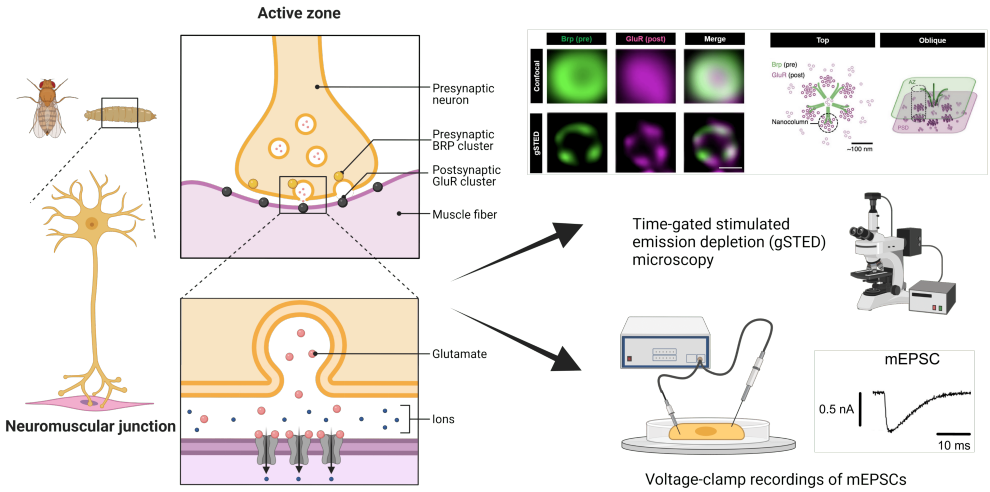


Figure 7.1: **Description of the experimental apparatus**. Observations were obtained from the Drosophila neuromuscular junction (left). Two kinds of observations are used: firstly, **microscopy observations** obtained using gSTED microscopy (upper right). It allows to determine, for several active zones, the position of presynaptic BRP proteins (which are used as a proxy for the position of release sites) and of postsynaptic GluR clusters. Secondly, **electrophysiology observations** obtained using voltage-clamp (lower right). It allows to record mEPSCs due to the spontaneous release of vesicles at active zones.

functions and pairwise interactions, and perform model selection to compare their goodness-of-fit. Finding out which model better fits the data allows to draw conclusions on how clusters are distributed and interact. Unless specified otherwise, analyses were performed using the `R` package `spatstat` [211, 212].

A point process with specified pairwise interactions is called a Gibbs Point Process (GPP). The probability distribution for the positions of a set of observations $\mathbf{X}, \mathbf{Y}$ is computed as

$$p(\mathbf{X},\mathbf{Y}) \propto \left(\prod_{i=1}^{n} b^X(\mathbf{x}^{(i)})\right)\left(\prod_{j=1}^{m} b^Y(\mathbf{y}^{(j)})\right)\left(\prod_{i<i'}^{n} c^{X,X}(\mathbf{x}^{(i)},\mathbf{x}^{(i')})\right) \left(\prod_{j<j'}^{m} c^{Y,Y}(\mathbf{y}^{(j)},\mathbf{y}^{(j')})\right)\left(\prod_{i,j}^{n,m} c^{X,Y}(\mathbf{x}^{(i)},\mathbf{y}^{(j)})\right) \quad (7.1)$$

where $\mathbf{X} = \{\mathbf{x}^{(i)}\}_{i=1}^{n}$, $\mathbf{x}^{(i)} = (x_1^i, x_2^i)$ are the 2D coordinates of n BRP clusters and $\mathbf{Y} = \{\mathbf{y}^{(j)}\}_{j=1}^{m}$, $\mathbf{y}^{(j)} = (y_1^j, y_2^j)$ are the 2D coordinates of m GluR clusters.

First, point processes are characterized by their intensity function $b^z(\mathbf{u})$ (i.e. the density of points), which can be either uniform or be a function of the position $\mathbf{u} = (x_1, x_2)$ in the 2D plane and of the type z of points, with $z \in \{X, Y\}$. We propose 3 different intensity functions:

- **Uniform**: $\log b^z(x_1, x_2) = -\theta_z$

Under the uniform hypothesis, the process is said to be homogeneous (i.e. the point density is the same over the whole surface of the AZ).

- **Distance to center**: $\log b^z(x_1, x_2) = -\theta_z \sqrt{x_1^2 + x_2^2}$

  Under this hypothesis, the density of proteins is assumed to be a decreasing function of the distance to the center of the AZ. Different distance functions, e.g. a Gaussian kernel $\exp\left(-\frac{x_1^2+x_2^2}{2c^2}\right)$, can also be investigated.

- **Ring patterns**: $\log b^z(x_1, x_2) = -\theta_z |\sqrt{x_1^2 + x_2^2} - r_z|$

  where r_z is the mean distance of proteins of type z to the center of the AZ. Indeed, BRP and GluR have been observed to be organized as ring-like patterns, with an inner ring of BRP proteins close to the center of the AZ and an outer ring of GluR clusters (Figure 7.1, upper right panel). Performing model selection would precisely allow to verify this hypothesis, by confronting its goodness-of-fit to that of other models.

The value of the free parameter θ_z is computed using the Berman-Turner algorithm [213], as detailed in Section 11.3 in [212].

Second, point processes are characterized by the interpoint interactions, i.e. by possible dependencies between the points' positions and types. The functions $c^{z,z'}$ correspond to pairwise interaction terms, which are symmetric between positions $\mathbf{u}$ and $\mathbf{v}$ (i.e. $c^{z,z'}(\mathbf{u}, \mathbf{v}) = c^{z,z'}(\mathbf{v}, \mathbf{u})$). Here, we propose 4 different interaction schemes:

- **Poisson point process**: $c^{z,z'}(\mathbf{u}, \mathbf{v}) = 1 \ \forall (z, z') \in \{X, Y\}$

  In a Poisson point process, all points are i.i.d., irrespective of their types.

- **Hardcore point process**: $c^{z,z'}(\mathbf{u}, \mathbf{v}) = \begin{cases} 1 & \text{if } ||\mathbf{u} - \mathbf{v}|| > r \\ 0 & \text{if } ||\mathbf{u} - \mathbf{v}|| \leq r \end{cases} \forall (z, z') \in \{X, Y\}$

  A Hardcore process specifies a minimal distance r between two points $\mathbf{u}$ and $\mathbf{v}$.

- **Strauss point process**: $c^{z,z'}(\mathbf{u}, \mathbf{v}) = \begin{cases} 1 & \text{if } ||\mathbf{u} - \mathbf{v}|| > r \\ \gamma & \text{if } ||\mathbf{u} - \mathbf{v}|| \leq r \end{cases} \forall (z, z') \in \{X, Y\}$

  A Strauss process is a generalization of the previous processes. It reduces to a Poisson process for $\gamma = 1$ and to a Hardcore process for $\gamma = 0$.

- **Multi-Strauss point process**[2]: $c^{z,z'}(\mathbf{u}, \mathbf{v}) = \begin{cases} 1 & \text{if } ||\mathbf{u} - \mathbf{v}|| > r_{z,z'} \\ \gamma_{z,z'} & \text{if } ||\mathbf{u} - \mathbf{v}|| \leq r_{z,z'} \end{cases} \forall (z, z') \in \{X, Y\}$

  It further generalizes the Strauss process to be type-dependent: the probability parameter γ and the interaction radius r will depend on the respective types of $\mathbf{u}$ and $\mathbf{v}$ (*BRP-to-BRP*, *BRP-to-GluR*, or *GluR-to-GluR*).

These 3 intensity functions and 4 pairwise interaction functions thus define 12 possible generative models (Figure 7.2), which are compared based on their respective AIC for the obtained microscopy observations (Figure 7.3, see Section 1.6 for the derivation of the AIC). For WT subjects, the model that has the best AIC has an intensity function structured in ring-like patterns. For mutants, it is a model in which the

[2]Contrary to a Strauss point process, in a Multi-Strauss, under some conditions the interaction term $\gamma_{z,z'}$ can be greater than 1 (see [212], section 25.1.4).

intensity scales with the distance to the center. We can thus conclude that clusters are more organized (in ring patterns) in WT than in mutants.

For WT subjects, the best model of interaction is a simple Strauss process (independent from the types), while mutants data are better fitted with a Multi-Strauss process. For this Multi-Strauss process the gamma parameter for the *GluR-to-BRP* interaction verifies $\gamma_{BRP,GluR} > 1$. We can thus conclude that clusters are not Poisson, and that in mutants GluR and BRP have a tendency to lie close to each other.

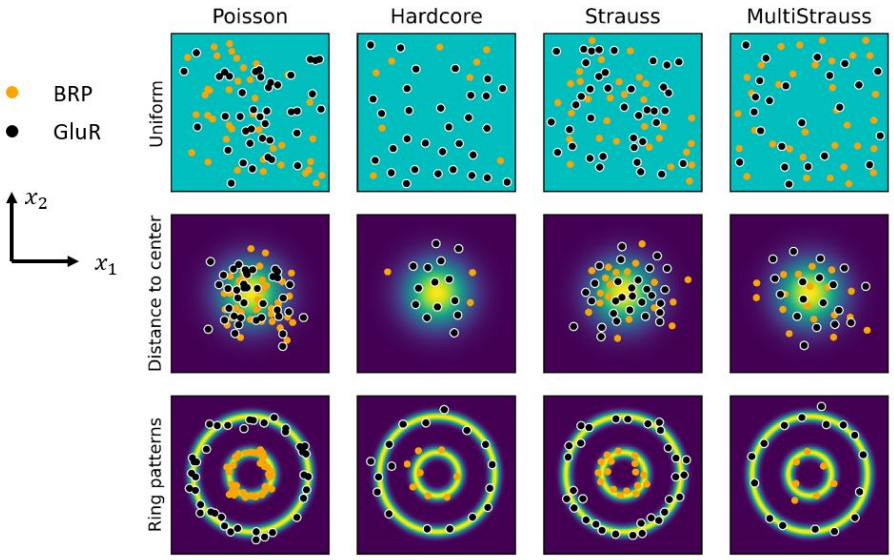


Figure 7.2: **Illustrations of the 12 different possible models of point patterns**. A model is defined by its intensity function (*Uniform*, *Distance to center*, or *Ring patterns*), and by its pairwise interaction function (*Poisson*, *Hardcore*, *Strauss*, or *Multi-Strauss*).

We thus have obtained generative models (one for WT subjects, and one for Unc13A mutants) for the position and interaction of synaptic proteins.

7.4 A simple model of GluR activation

A second step is to design a model of glutamate receptor sensitivity, i.e. to compute the probability that a GluR will be activated as a function of its distance to a neurotransmitter release site. The process of mEPSC generation following a spontaneous presynaptic release involves three main steps:

1. **The release of glutamate at the presynaptic side**: in spontaneous release, a presynaptic vesicle fuses with the plasma membrane and releases its content of neurotransmitter molecules in the synaptic cleft. This release process can be approximated as being instantaneous, or can be modelled as an exponentially decaying source of glutamate.

2. **The spread of glutamate in the synaptic cleft**: this diffusion corresponds to a random walk of glutamate molecules. Given the small cleft width (compared to the surface of the AZ), it is often approximated as a 2-dimensional diffusion process, with an experimentally observed diffusion coefficient.

3. **The activation of postsynaptic GluR**: finally, these glutamate molecules

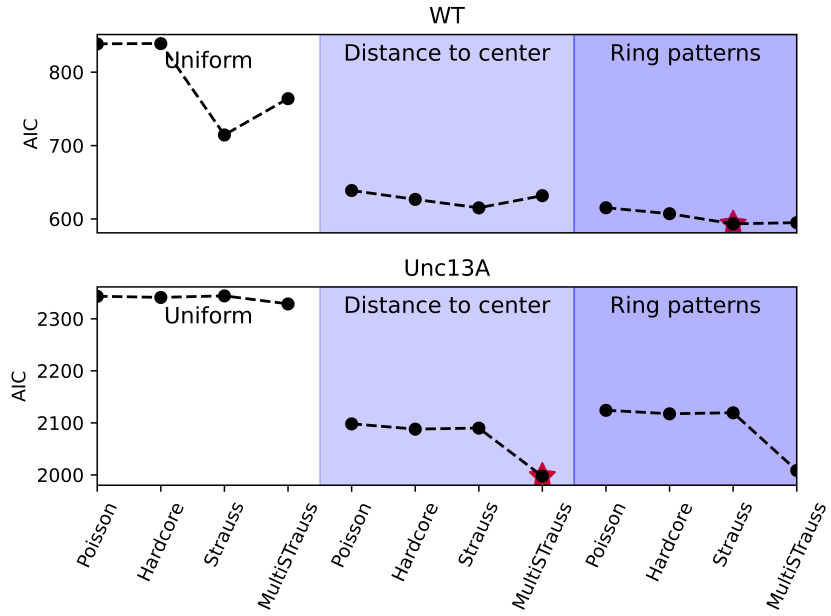


Figure 7.3: **Results of model fitting**. A lower AIC indicates a better fit. For WT subjects, the model that has the best AIC has an intensity function structured in ring-like patterns, and the best model of interaction is a simple Strauss process (independent from the types). For mutants, it is a model in which the intensity scales with the distance to the center, and where the interaction function is a Multi-Strauss process where the gamma parameter for the *GluR-to-BRP* interaction verifies $\gamma_{BRP,GluR} > 1$.

will bind to postsynaptic receptors and activate them, thus eliciting a postsynaptic current. GluRs are organized in clusters of approximately 20 receptors, each of them being independent (i.e. a cluster does not activate in an all-or-nothing fashion). GluRs are complex biophysical structures, which are often modelled as kinetic state diagrams (e.g. Figure 7.4). For instance, in [214], the authors explain that "*AMPAR kinetics were modeled using [...] a 7 state model and published values for kinetics rate constants, adjusted to fit the experimentally recorded mEPSCs in our system. In this scheme, resting AMPARs are in a C1 state and move to states C2 and C3 upon successive binding of two glutamate molecules. O4 is the open state, whereas D5, D6, and D7 are desensitized states. Mono-liganded D7 desensitized AMPARs can move to the biliganded D6 state.*" Such pharmacodynamical models nicely fit recorded mEPSCs [214] but rely on several kinetics rate constants, which may be hard to accurately estimate from data and prevent the model from correctly generalizing to future observations.

Different models of glutamate release and GluR activation are summarized in Table 7.1. All models are based on CNS synapses (hippocampal neuron cultures and simulations of CA1 cells in [215], CA1 pyramidal cells in [216], cultured rat hippocampal neurons in [217], corticogeniculate and retinogeniculate synapses in the dorsallateral geniculate nucleus in [218]). They all assume 2000 molecules of glutamate per vesicle release (between 2500 and 5000 in [218]), and a cleft width of $15nm$. Interestingly, all of them (except for [217]) assume a complete pharmacodynamical model of postsynaptic receptors (as in [214] and Figure 7.4): for these models, the activation probability of a receptor as a function of its distance to the glutamate source can

Paper	Uptake	Receptor	Release	Diffusion
[215]	×	Complete pharmacodynamical model	Not mentioned	Not mentioned
[216]	✓	Complete pharmacodynamical model	Pore diffusion	2D diffusion process ($0.4\mu m^2/ms$)
[217]	×	Simplified model	NA	NA
[219]	×	Complete pharmacodynamical model	Instantaneous (same model as in [218])	2D diffusion process ($0.3\mu m^2/ms$)
[218]	✓	Simplified pharmacodynamical model (2 states)	Either instantaneous or exponentially decaying source of glutamate	2D diffusion process ($0.3\mu m^2/ms$)

Table 7.1: **Review of models of glutamate release and GluR activation. Uptake** indicates whether glutamate uptake is modelled. **Receptor** refers to how postsynaptic receptors are modelled: either using a complete pharmacodynamical model as in Figure 7.4, or using a simplified model with less free parameters. **Release** refers to how neurotransmitter molecules release at the fusion pore is modelled. **Diffusion** refers to how the random walk of glutamate molecules is modelled.

only be computed via lengthy numerical simulations. Our goal here is thus to derive a simpler model of GluR activation, i.e. to define the probability that a receptor will be activated as a function of its distance to the release site.

Given that the depth of the synaptic cleft is small compared to the total surface of the AZ, we use the "expanding cylinder approximation" to represent the spread of glutamate: after the fusion of a presynaptic vesicle with the plasma membrane, the volume occupied by glutamate is assumed to be a cylinder which basis has radius r and is centered on the release site, and which height corresponds to the depth of the synaptic cleft. Assuming that a spontaneous release happened at one of the presynaptic release sites, we use the following notations:

- r is the distance from the release site up to which glutamate has spread [nm];
- z is the depth of the synaptic cleft [nm];
- n is the number of released glutamate molecules [-];
- m is the number of these molecules required to activate one receptor [-];
- ΔV is the volume of a voxel corresponding to the surface of a postsynaptic receptor and to the depth of the synaptic cleft [nm^3];
- L is the number of these voxels into the cylinder of radius r (i.e. covered by glutamate), i.e. $L = \frac{\pi r^2 z}{\Delta V}$;
- ρ is the glutamate density defined as $\rho = \frac{n}{L} = \frac{n\Delta V}{\pi r^2 z}$

We firstly assume that $m = 1$ and that, among L voxels, one contains a receptor. The probability $p_{\text{open}}(n, L)$ that this receptor will be activated can be approximated as the probability that at least one of the n molecules will be in the same voxel as

the receptor:

$$p_{\text{open}}(n, L) = 1 - \left(1 - \frac{1}{L}\right)^n \tag{7.2}$$

Using the current value of the density ρ yields

$$p_{\text{open}}(n, \frac{n}{\rho}) = 1 - \left(1 - \frac{\rho}{n}\right)^n \tag{7.3}$$

Finally, assuming that n and L are large (keeping ρ fixed) yields

$$\lim_{n\to\infty} p_{\text{open}}(n, \frac{n}{\rho}) \approx 1 - e^{-\rho} = 1 - e^{\frac{-\nu_0^2}{r^2}} \tag{7.4}$$

with $\nu_0 = \sqrt{\frac{n\Delta V}{\pi z}}$.

For $m > 1$, Eq. 7.2 becomes

$$p_{\text{open}}(n, L) = \sum_{k=m}^{n} \binom{n}{k} \left(\frac{1}{L}\right)^k \left(1 - \frac{1}{L}\right)^{n-k} = 1 - \sum_{k=0}^{m-1} \binom{n}{k} \left(\frac{1}{L}\right)^k \left(1 - \frac{1}{L}\right)^{n-k} \tag{7.5}$$

which, for large n and L and keeping ρ fixed, becomes

$$p_{\text{open}}(n, L) \approx 1 - e^{-\rho} \sum_{k=0}^{m-1} \frac{\rho^k}{k!} \tag{7.6}$$

which is equal to 0 when m converges to infinity. The probability that a postsynaptic receptor is activated by a glutamate release at a distance r can thus be approximated by $1 - e^{-\nu_0^2/r^2}$, where ν_0 is a parameter depending on the geometry of the synaptic cleft.

However, the physical proximity between a glutamate molecule and a postsynaptic channel does not necessarily imply that the latter will be opened. Their molecular affinity is not perfect, and a receptor needs to be in a given state to be activated by a glutamate molecule. Using the random variables $X \in \{0, 1\}$ to represent the presence of a glutamate molecule at a receptor position, and $S \in \{0, 1\}$ to represent the state of the said receptor, its activation probability will be $p(X = 1)p(S = 1)$. Using the notations $p(X = 1) = p_{\text{open}}(n, \frac{n}{\rho})$ (Eq. 7.4) and $p(S = 1) = 1/\nu_1$ with $1 < \nu_1$, our proposed model of probability of channel opening as a function of the distance between the vesicle release site and the receptor is finally

$$p(r) = \frac{1}{\nu_1}(1 - e^{-\nu_0^2/r^2}) \tag{7.7}$$

where:

- ν_0 is a shape parameter defining the slope of the function, i.e. how fast the activation probability will decrease with the distance r;
- ν_1 is a scale parameter defining the affinity between glutamate and postsynaptic receptors (as a distance $r = 0$ does not necessarily imply that the postsynaptic receptor will be in an activated state).

To assess the validity of our model, we need to fit the parameters ν_0 and ν_1 to experimental data. A pharmacodynamical model of AMPA receptors activation (illustrated in Figure 7.4), which nicely fits recorded mEPSCs, is used in [214]. Using this model, the authors compute the number of open receptors in three different scenarii:

- **First scenario**: simulation of one GluR cluster and of one release site. The number of activated receptors is computed as a function of the distance between the release site and the cluster (Figure 7.5, left).
- **Second scenario**: simulation of one release site and of two clusters, one being located underneath the release site and the second one at a varying distance. The number of activated receptors is computed as a function of this distance (Figure 7.5, center).
- **Third scenario**: simulation of one release site flanked by two clusters. The number of activated receptors is computed as a function of the displacement of the release site (Figure 7.5, right).

The results of numerical simulations based on the pharmacodynamical model in [214] are plotted as orange lines on Figure 7.5. Results from our simplified model (Eq. 7.7) after fitting of ν_0 and ν_1 on data from [214] are plotted as blue lines. We thus verify that our function reproduces previous results on GluR sensitivity, without the need for long simulation of individual receptors dynamics.

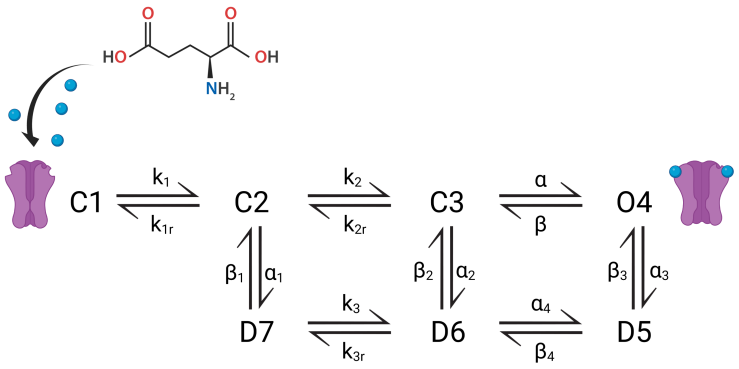


Figure 7.4: **Classical pharmacodynamical model of GluR activation**. Illustration of the model used in [214]: postsynaptic receptors are modelled using "*a 7 state model and published values for kinetics rate constants, adjusted to fit the experimentally recorded mEPSCs in our system. In this scheme, resting AMPARs are in a C1 state and move to states C2 and C3 upon successive binding of two glutamate molecules. O4 is the open state, whereas D5, D6, and D7 are desensitized states. Mono-liganded D7 desensitized AMPARs can move to the biliganded D6 state.*"

7.5 Parameter fitting to mEPSC data

The third step is to link our proposed model of GluR sensitivity (Eq. 7.7) to the observed distribution of recorded mEPSCs, which were recorded via voltage-clamp (Figure 7.1, lower right). We propose the following generative model for mEPSCs:

$r_h^{(i,j)}$ is the distance between release site $\mathbf{x}_h^{(i)}$ and postsynaptic receptor $\mathbf{y}_h^{(j)}$ at the h-th Active Zone of the studied NMJ:

$$r_h^{(i,j)} = ||\mathbf{x}_h^{(i)} - \mathbf{y}_h^{(j)}|| \tag{7.8}$$

$k_h^{(i,j)}$ is the number of receptors in cluster $\mathbf{y}_h^{(j)}$ activated by a release at site $\mathbf{x}_h^{(i)}$:

$$k_h^{(i,j)} | r_h^{(i,j)} \sim \text{Bin}(p(r_h^{(i,j)}), N) \tag{7.9}$$

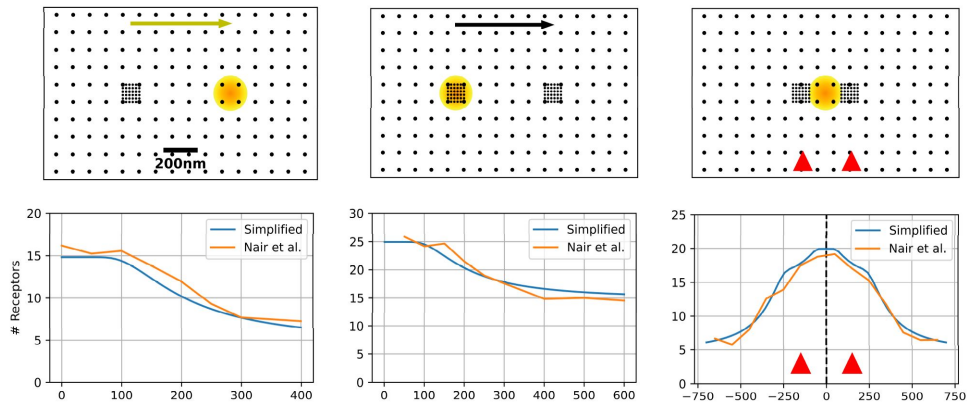


Figure 7.5: **A simple model of GluR activation**. Left: 1 BRP, 1 GluR, different positions of the BRP. Center: 1 BRP, 2 GluR, different positions of one of the GluR. Right: 1 BRP, 2 GluR, different positions of the BRP. Orange lines: results from simulations using the complete pharmacodynamical model in [214]. Blue lines: results computed analytically using our model (Eq. 7.7). Parameter values of the simplified model after fitting on the data from [214]: $\nu_0 = 177.24$ [nm] , $\nu_1 = 2.47$.

where N is the number of GluR per cluster, and $p(\cdot)$ is the sensitivity function defined by Eq. 7.7; when a spontaneous release event happens at site $\mathbf{x}_h^{(i)}$, mEPSC $e_h^{(i)}$ is thus defined as

$$e_h^{(i)} = q_0 \sum_{j=1}^{j_h} k_h^{(i,j)} + \epsilon \tag{7.10}$$

where j_h is the number of GluR clusters at AZ number h, $\epsilon \sim \mathcal{N}(0, \sigma^2)$ represents the recording noise with variance σ^2, and q_0 is the postsynaptic current elicited by one receptor activation.

The distribution of $e_h^{(i)}$ is then approximated as a normal distribution, i.e. $e_h^{(i)} \sim \mathcal{N}(\mu_{e_h^{(i)}}, \sigma^2_{e_h^{(i)}})$, with

$$\mu_{e_h^{(i)}} = q_0 \cdot N \cdot \sum_{j=1}^{j_h} p(r_h^{(i,j)})$$

and

$$\sigma^2_{e_h^{(i)}} = \sigma^2 + q_0^2 \cdot N \cdot \sum_{j=1}^{j_h} p(r_h^{(i,j)})(1 - p(r_h^{(i,j)}))$$

Using the function `generate` from the `R` package `spatstat`, $H = 200$ independent surrogate AZ were generated from the fitted model described in Section 7.3. Assuming that AZ number $h \leq H$ has i_h release sites and j_h postsynaptic receptor clusters, this allows to obtain empirical samples $r_h^{(i,j)}$ (with $1 \leq h \leq H$, $1 \leq i \leq i_h$, and $1 \leq j \leq j_h$) for the distances between release sites and receptors (Eq. 7.8). The final probability density function for mEPSCs is finally computed as

$$p(e) = \frac{1}{H} \cdot \sum_{h=1}^{H} \frac{1}{i_h} \cdot \sum_{i=1}^{i_h} p(e_h^{(i)}) \tag{7.11}$$

	Recording 1	**Recording 2**	**Recording 3**
N [-]	20	20	20
ν_0 [nm]	142.39	125.84	168.02
$1/\nu_1$ [-]	0.11	0.10	0.14
q_0 [nA]	0.13	0.163	0.07
σ [nA]	$7.52 \cdot 10^{-5}$	$6.94 \cdot 10^{-6}$	$5.27 \cdot 10^{-12}$

Table 7.2: **Results of model fitting on mEPSCs recordings.** Based on the spatial distribution of BRP and GluR clusters described in Section 7.3, the values of the parameters of our generative model of spontaneous mEPSCs (Equations 7.7 to 7.10) were fitted on 3 different sets of mEPSCs, containing respectively 220 (**Recording 1**), 266 (**Recording 2**), and 263 (**Recording 3**) data points.

This distribution is fitted on the observed empirical distribution of mEPSCs to infer the values of parameters N, ν_0, ν_1, q_0, and σ (see results in Table 7.2 and Figure 7.6).

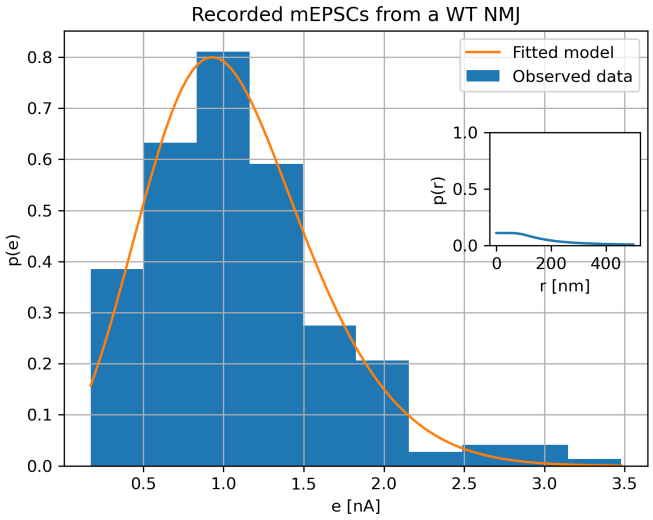


Figure 7.6: **Results of model fitting on mEPSCs recordings (Recording 1).** Blue: histogram of the recorded mEPSCs in **Recording 1** (220 data points). Orange: probability distribution of mEPSCs (Eq. 7.11) which parameters have been fitted (see values in Table 7.2) to the observed distributions of mEPSCs. Inset: value of $p(r)$ (Eq. 7.7) for the fitted values of ν_0 and ν_1.

7.6 Modeling synaptic transmission predicts that release sites locate within GluR nano-rings for effective GluR activation

Our fourth and last step is to derive a normative prediction of the position of release sites. We propose to compute the optimal position for the release sites from an information transmission argument, and to assess how it compares to the observed position of BRP clusters.

It follows from gSTED observations that WT AZ have on average 14 GluR clusters

on the postsynaptic density, and that their mean distance to the center of the AZ is 225nm. An average WT AZ (i.e. $h = 1$) is thus created by modeling 14 GluR clusters on a ring structure with radius 225nm around the center of the AZ (black dots and red circle in Figure 7.7). We then assess how two metrics (the mean mEPSC μ_e and the SNR μ_e/σ_e) evolve as a function of the position of a single synthetic release site (i.e. $i_h = 1$).

The mean mEPSC μ_e is computed as

$$\mu_e = q_0 \cdot N \cdot \sum_{j=1}^{14} p(r_1^{(1,j)}) \tag{7.12}$$

while its variance is equal to

$${\sigma_e}^2 = \sigma^2 + q_0^2 \cdot N \cdot \sum_{j=1}^{14} p(r_1^{(1,j)})(1 - p(r_1^{(1,j)})) \tag{7.13}$$

which allows to compute the SNR μ_e/σ_e. In Equations 7.12 and 7.13, the term $r_1^{(1,j)}$ refers to the distance between the position of the considered release site and the j-th GluR cluster. The value of the parameters N, ν_0, ν_1, q_0, and σ, on which the values of μ_e and μ_e/σ_e depend, are fitted on experimental data (Table 7.2), as explained in Section 7.5.

In Figure 7.7 (left panel), the blue color map indicates the mean mEPSC elicited by releasing glutamate at different points of the AZ and computed from Eq. 7.12. The dashed yellow line shows its maximum while the solid yellow line corresponds to the average distance between observed release sites and the center of the AZ. The right panel shows similar results but for the SNR. In both cases, preliminary results show that the observed position of BRP proteins is well in line with the position predicted from a normative argument (Figure 7.8).

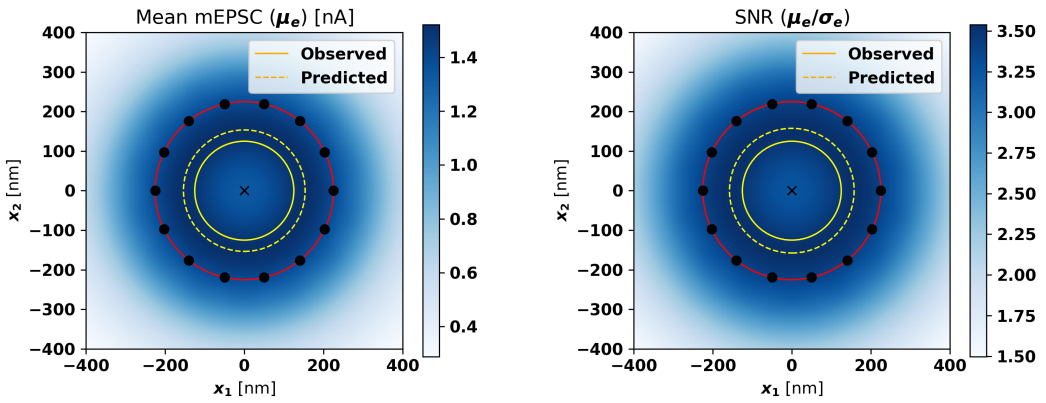
It is interesting to see that two regimes appear depending on the value of ν_0 (Figure 7.9): at low ν_0 (low glutamate spread), it is optimal for the release sites to be close to the postsynaptic clusters; at high ν_0 (high glutamate spread), it is optimal to be close to the center of the AZ to activate several postsynaptic clusters.

7.7 Conclusion

Our preliminary contribution is three-fold. Firstly, we derived a generative model for the organization of synaptic proteins, which accounts for both their position on the AZ and the intercations between presynaptic BRP and postsynaptic GluR. Model comparison supports ring-like patterns for BRP and GluR clusters at the Drosophila NMJ, which are disrupted in Unc13A mutants. Secondly, we derived a simpler model of glutamate spread and GluR activation, which nicely fits previous results without the need for lengthy numerical simulations or a large number of free parameters. Finally, modeling synaptic transmission after systematic variation of release site location predicts that release sites locate within GluR nano-rings for effective GluR activation.

These are preliminary results that need to be consolidated:

- More microscopy observations, and observations of different presynaptic proteins, are required, as BRP proteins are only loose proxies for the actual position of release sites;



(a) (b)

Figure 7.7: **A normative approach to the organization of synaptic proteins (1)**. (a) The blue color map indicates the mean mEPSC elicited by releasing glutamate at different points of the AZ for the parameters values for **Recording 1** (Table 7.2). Black dots represent GluR clusters. The dashed yellow line shows its maximum while the solid yellow line corresponds to the average distance between observed release sites and the center of the AZ. To maximize the mean mEPSC, it is optimal for the release sites to be located on a ring structure between the VGCC and the GluR clusters. This theoretical prediction is well in line with microscopy's observations. (b) Similar results for the SNR.

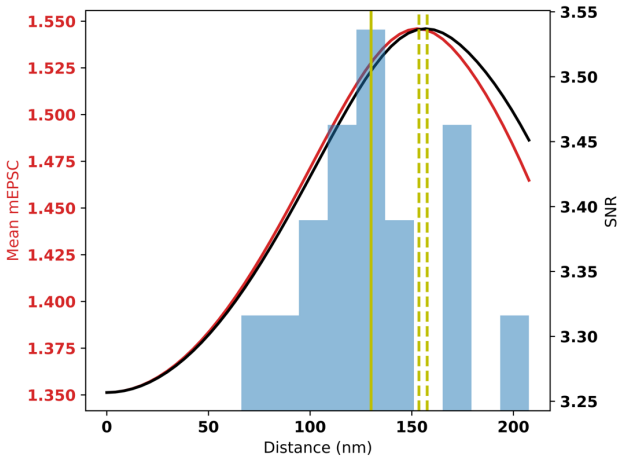


Figure 7.8: **A normative approach to the organization of synaptic proteins (2)**. Blue histogram: distribution of the observed distances between BRP proteins and the center of the AZ (the solid yellow line corresponds to its mean). Red line: mean mESPC μ_e as a function of the distance between the release site and the center of the AZ, computed from Eq. 7.12. Black line: SNR μ_e/σ_e as a function of the distance between the release site and the center of the AZ, computed from Eq. 7.13. The dashed vertical yellow lines correspond to their respective maxima.

- Similarly, results need to be consolidated and tested on more electrophysiology recordings;

- Most statistical analyses were performed using the `R` package `spatstat`, in which the parameters of several built-in functions (e.g. the correction for boundary effects when plotting Ripley's K-function, or the method used to generate samples from a point process model) are not easily modifiable by the user. Repeating these analyses using hand-made routines is a work in process, which would allow to robustify the results.

- Finally, when fitting the value of the parameters of our model on observed mEPSCs (in Section 7.5), biological constraints must be defined in order to ensure that the fitted values of N, ν_0, ν_1, q_0, and σ are within realistic ranges. This is all the more important as the fitted values of some parameters, and especially of ν_0 and ν_1, have been observed to be correlated. Moreover, results in Table 7.2 indicate 6 orders of magnitude between the fitted values of σ in Recording 2 and in Recording 3, while these fitted values have not been compared to the expected recording noise from the recording apparatus. This inference step thus needs to be consolidated to ensure the validity of the normative approach presented in Section 7.6. Especially, it would be interesting to perform a sensitivity analysis, to assess how the results illustrated in Figure 7.7 are robust to a change in the value of the fitted parameters.

Many recent machine-learning methods consider uncertainty. For example, Bayesian neural networks rely on the posterior distribution of synaptic weights given all the observed data. However, it is largely unclear how synapses represent this distribution. Based on the quantal theory of synaptic transmission, the smallest unit of synaptic information transfer is the postsynaptic response to the fusion of individual synaptic vesicles, which is typically assayed by the amplitude of spontaneous miniature exci-

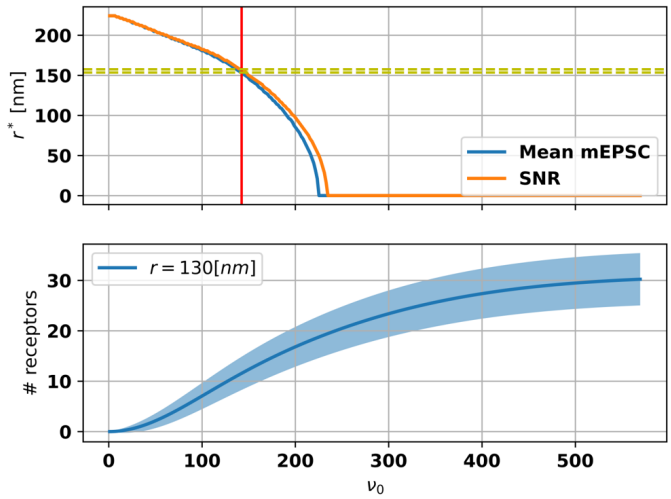


Figure 7.9: **Optimal position for the release sites as a function of the shape parameter** ν_0. Upper panel: optimal distance r^* between the release site and the center of the AZ to maximize either μ_e (blue line) or the SNR (orange line). Yellow dashed lines correspond to the optima displayed in Figure 7.7. The vertical red line corresponds to the fitted value of ν_0 for **Recording 1**. Two regimes appear: at low ν_0 (low glutamate spread), it is optimal for the release sites to be close to the postsynaptic clusters; at high ν_0 (high glutamate spread), it is optimal to be close to the center of the AZ to activate several postsynaptic clusters. Lower plot: mean number of activated receptors as a function of ν_0 for a distance r=130nm between the release site and the center of the AZ.

tatory postsynaptic currents (mEPSCs). To investigate the underlying mechanisms of mEPSCs variability, we modeled synaptic transmission based on the sub-synaptic distribution of Glutamate Receptors (GluRs) clusters at the Drosophila NMJ revealed by Time-Gated Stimulated Emission Depletion (gSTED) microscopy. Our model of spontaneous synaptic transmission will be critical for the validation of normative theories of Bayesian synapses, and to reveal how synaptic weight distributions or uncertainties are represented at the synaptic level.

Chapter 8

Discussion and perspectives

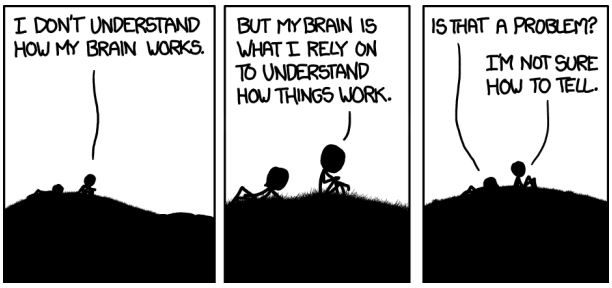


Figure 8.1: On studying how the brain works[1].

8.1 Summary of results

Synaptic characterization, i.e. the methods used to describe a synapse, to derive a generative model of it, and to infer its parameters, is a field of crucial importance for both experimental and theoretical neuroscientists. However, current experimental techniques and analysis tools suffer from several drawbacks: the goal of this doctoral thesis was to propose solutions to address them.

Firstly, methods that are still commonly used by experimentalists to infer the parameters of a synaptic connection, such as mean-variance analysis, or cumulative EPSC methods, have important pitfalls. For instance, mean-variance analysis requires to perform recordings under different calcium concentrations, and, for each of these settings, to obtain enough data points to accurately estimate the mean and the variance: obtained estimates are thus often highly unreliable. Similarly, cumulative EPSC methods are only applicable to depressing synapses and cannot be used with arbitrary patterns of activation. In Chapter 3, we reviewed more promising model-based approaches, which can be applied with any stimulation protocols, and can be used to infer both synaptic release statistics and synaptic dynamics. We expect these probabilistic inference methods to help experimentalists to obtain more accurate estimates of synaptic transmission parameters.

Although promising, these model-based approaches require to posit a generative model $\mathcal{M}$, which needs to have basic statistical properties. An important property of a model is called identifiability: a model is said to be structurally identifiable if

[1] `https://xkcd.com/1163/`, use with authorization.

its parameters can be uniquely inferred from the distribution of its outputs. This theoretical property has been thoroughly defined and characterized, but practical identifiability, which is a property of both the model and the experimental protocol, is usually only qualitatively assessed. In Chapter 4, we proposed a statistical criterion for practical identifiability, which allows experimentalists to answer crucial questions about their experiments: for instance, whether the recording noise is sufficiently low to ensure peaks and dips in the distribution of observations are visible, or if presynaptic stimulation times are sufficiently informative.

Moreover, using a model-based approach with a certain generative model $\mathcal{M}$ yields two important issues. Firstly, it requires the inputs to the model (i.e. the timings of the presynaptic action potentials) to be sufficiently informative about the parameters, and to be adequately designed. Secondly, the computation of the likelihood function (which is used either to obtain a point-based estimate of the parameters, to compute their posterior distribution, or to update the samples in a particle filtering scheme) can be challenging. For instance, when representing a synapse with a state-space model, the computation time scales polynomially with the number of vesicles. We address both issues in Chapter 5: we proposed an online and parallel particle filtering scheme for efficient computation of the posterior distribution of the parameters, and design a closed-loop inference scheme to compute the next stimulation time which is maximally informative about the hidden parameters. Besides, in Appendix A, we derive a mean-field approach to simplify the computation of the likelihood.

An issue that is too often disregarded by experimentalists is the question of model comparison and selection. To ensure that synaptic parameters are precisely estimated, the generative model $\mathcal{M}$ needs to be an accurate description of the studied synapse, and to avoid both under- and overfitting. In Chapter 6, we suggest to use our proposed definition of practical identifiability to perform a form of Optimal Experiment Design, and to optimally compare an alternative model $\mathcal{M}_1$ to a null model $\mathcal{M}_0$. Besides, in Appendix B, we discuss the possibility to use the BIC (a classically used model selection criterion) to models with correlated data. This is especially useful for synaptic characterization, as synapses are commonly described using Markov processes, in which observations are not i.i.d. Overall, we expect these results to generalize the use of model selection tools in synaptic characterization.

Finally, we apply the different theoretical tools used in the previous chapters (model selection, model fitting, and mean-field approach) to study synapses at a different scale and to make sense of the observed organization of synaptic proteins at individual active zones. In Chapter 7, we firstly use model selection to derive a generative model of the organization of presynaptic BRP proteins and postsynaptic GluR clusters: this illustrate that model selection can be used as a form of hypothesis testing. Then, using a mean-field approach, we derive a model of GluR sensitivity to glutamate release, which allows to model postsynaptic currents more efficiently than previously used pharmacodynamical models. The free parameters of this model are inferred using model fitting on recorded mEPSCs; finally, we propose a normative approach to the structure of active zones, and compare our theoretical predictions to empirical results.

Overall, we expect the results presented in this doctoral work to improve the way synaptic characterization is being performed.

8.2 Related work

The different techniques and tools that we proposed in this thesis are building up on previous work, and aimed at improving existing methods. For instance, our proposed definition of practical identifiability (Chapter 4) is an alternative to previous definitions (see for instance [70, 72, 189]), which are simple and intuitive but are data-dependent instead of defining an intrinsic property of a model. The identifiability domains we plot correspond to an extension of the *landscaping* technique described in [153], in which the possibility to discriminate two competing models is assessed by measuring their respective evidences on a large number of surrogate data sets. We extend this technique by averaging the model evidences over the distribution of data sets. We also extend the *expected support of a model* introduced in [154] to several models, comparing their respective supports to define a criterion for identifiability. Our theoretical considerations on model identifiability (described in Appendix C) can also be seen as a more general version of a result presented in the following paper using less stringent approximations [65].

Parameter estimates from model-based approaches are often associated with high uncertainty (see for instance [20] or [26]). To correct this pitfall, previous studies had already pointed out the need to optimize the presynaptic stimulation protocol to maximize its informativeness. Previous work showed that some stimulation protocols are more informative than others, but ignored the temporal correlations of the number of readily-releasable vesicles [159] or did not compute which protocol would be most informative [20]. The idea of a closed-loop optimization scheme to compute the next stimulation time was originally proposed in our review on model-based approaches (see [3] and Chapter 3): "*Pushing this idea even further, it would be interesting to design a closed-loop inference scheme such that after each spike and its subsequent postsynaptic response, the algorithm determines the best interval for the next spike that is maximally informative about the synaptic parameters*". Our proposed solution, the ESB-BAL method described in Chapter 5, builds on previous applications of OED to neuroscience [29, 185, 184, 190] while being more general (it can be applied to any state-space model, even input-dependent ones) and time-efficient. Especially, variational approach [81] could be employed, but they usually require long computation times that can be impractical if the time between successive experiments is short. Closed-form solutions can only be computed for some special cases, such as linear models of GLM [29]. In particular, the particle filtering scheme that we use is an extension of the seminal Nested Particle Filter ([186]), which we extended to input-dependent state-space models.

Finally, in Chapter 4 and in [134], we discussed the possibility to extend the use of the BIC to correlated data sets. Indeed, the classical definition of the BIC is derived under the hypothesis that observations are i.i.d., which does not hold for general HMMs. To circumvent this pitfall, our approach in [134] was to analytically compute the Hessian matrix, but such a computation can be cumbersome. Posterior to our publication on practical identifiability, a very insightful paper, written by Shouto Yonekura, Alexandros Beskosa, and Sumeetpal S. Singh, was published in *Stochastic Processes and their Applications*. In this paper [220], the authors also address the validity of classically used Model Selection Criteria (including the BIC) for general Hidden Markov Models. Their main result is that, under regularity conditions (namely, that the observation generating process is strongly stationary and ergodic), the BIC remains strongly consistent for HMMs. In Appendix B, we build on this result and provide empirical evidence that it still holds for input-dependent HMMs,

given that the distribution of inputs is itself stationary. Providing a theoretical proof to this intuition is a work in progress.

8.3 Limitations

Our work has several limitations. Regarding our proposed definition of practical identifiability (Chapter 4), an obvious drawback compared to other definitions is that it requires to posit a nested null model $\mathcal{M}_0$. However, we argue that nested models and families naturally arise in commonly used statistical techniques, such as polynomial regression [148], or Generalized Linear Models (GLM) [160]. Especially, the widespread use of phenomenological models in neuroscience [161, 162, 163, 164] makes the use of nested families and submodels relevant.

Regarding our ESB-BAL framework, an important limitation is its variability over several repetitions, both in the estimates of parameters and in the computed optimal next input: indeed, it requires a very large number of particles to provide low variance estimates, as the approximation error only decreases with the square root of the number of particles. Also, we applied our method to a simplified example, where both the dimensions of the parameter space and of the possible experimental designs is relatively low (especially compared to experiments commonly performed in biology or in chemical engineering, for instance).

More generally, an obvious limitation of the statistical approaches to synaptic characterization that we proposed in this thesis is that they are applied to simplified models of synaptic transmission. Most of our application examples using surrogate data imply a simple binomial model of neurotransmitter release which does not account for important features, such as Short-Term Facilitation, variability in quantal amplitudes, existence of different pools of presynaptic vesicles, or postsynaptic receptors saturation and desensitization. This discrepancy between real synapses and our generative models might explain the results in Figure 5.5: theoretical results with simulated data show that using ESB-BAL significantly improves the accuracy of inferred parameters as compared to using a non-adaptative protocol, but results are less telling when applied to real synapses. This difference is likely due to the generative model being too simple and not a faithful representation of the synapse. Future work should thus focus on assessing whether our proposed approaches can scale to more complicated and realistic models.

8.4 Extensions and future work

In this concluding chapter, we highlight several possible applications for synaptic characterization, and discuss relevant future work.

8.4.1 Longitudinal study of synaptic transmission

A first application of our tools for synaptic characterization would be to infer synaptic parameters from a large number of synapses under various experimental conditions, and especially as a function of development stage. The output of this longitudinal study of synaptic transmission would be an open-access database containing both the electrophysiological recordings from several labs as well as the fitted model parameters obtained and compared via different methods (see Chapter 3). This database would be useful for both experimentalists and theoretical neuroscientists.

From the experimental point of view, applying different inference techniques on the same synapse may lead to variable results. This is all the more critical as synaptic parameters span a very broad range of different values, depending on the age of the subject (depressing synapses being more common in in young subjects, and facilitating ones being more common in adults [6]), on the concentrations of neurotransmitters (e.g. adenosine [221]), their localization in the CNS or in the PNS, or the type of neurons [7]. Having standardized inference techniques and reliable baseline estimates would be greatly beneficial to the field of synaptic characterization.

From the theoretical point of view, having access to parameters values from a large number of different synapse types would allow to derive and validate several predictions and normative theories concerning synaptic transmission. For instance, a burning question is to identify the computational relevant of stochasticity in transmission. A recent study, focused on geniculocortical neurons, shows that "*this variability and reliability actually enhance cortical firing via an increase in cortical membrane potential fluctuations, i.e., cortical firing rate is driven by fluctuations of Vm rather than mean Vm, similar to high conductance states*" [30]. Such claims could be compared to other recordings at different synapses showing a much higher reliability with almost no synaptic failures, and can be compared to different theories on the computational role of variability [39].

8.4.2 Different applications of ESB-BAL

In Chapter 5 we discussed how to use ESB-BAL for inferring the parameters of a model of chemical synapse. These parameters encompass both the binomial parameters (i.e. the number of vesicles N, their release probability p, and their quantal amplitude q), as well as the time constant τ_D characterizing STD. Here, we describe how to use ESB-BAL to specifically study the dynamics of vesicle refilling, and especially to study more complicated models of STD (instead of a complete model of neurotransmitter release).

As mentioned in the introduction, an important assumption of the classical binomial model is that release sites are similar and have the same homogeneous refilling time constant τ_D. However, it has been suggested [24] that presynaptic vesicles actually belong to two different groups: a readily releasable pool and a reluctantly releasable pool, having different refilling time scales. Assuming that the fast-refilling pool and the slow-refilling pool have respective refilling time constants $\tau_D^{(1)}$ and $\tau_D^{(2)}$, and that the fast-refilling pool makes up a fraction A_1 of the total number of vesicles, then the recovery (i.e. the evolution of the fraction of available vesicles $x(t)$ after presynaptic depletion) can be fitted with a biexponential trajectory [25]:

$$x(t) = A_1(1 - e^{-t/\tau_D^{(1)}}) + (1 - A_1)(1 - e^{-t/\tau_D^{(2)}}) \tag{8.1}$$

A classical way to assess the values of $\tau_D^{(1)}$ and $\tau_D^{(2)}$ is to first deplete the presynaptic pool of vesicles via sustained high-frequency stimulation, and then to probe the synapse after a chosen time interval Δ_t (recovery spike). The observed EPSC, expressed as a fraction of the first EPSC obtained before depletion, indicates the fraction of refilled vesicles after a time Δ_t.

In practice, after inducing depression via sustained high-frequency stimulation, only one recovery spike should be used to assess vesicles refill. Indeed, each presynaptic stimulation will induce a (possibly stochastic and ill-modelled) release of vesicles

which will perturb the refilling process. Using several recovery spikes would thus lead to a biased estimate of the refilling time constants. Given the limited time during which cells in acute brain slices can be recorded from, the number of possible observations (i.e. repetitions of the cycle sustained high-frequency stimulation + one recovery spike) is limited. ESB-BAL can thus be used to determine at which Δ_t the synapse should be probed to minimize the variance of the estimate of the refilling time constants, and to maximize the gain in information for a given number of observations.

Simulations in Figure 8.2 have been realized with the `OptimalVesicleSampling` package[2].

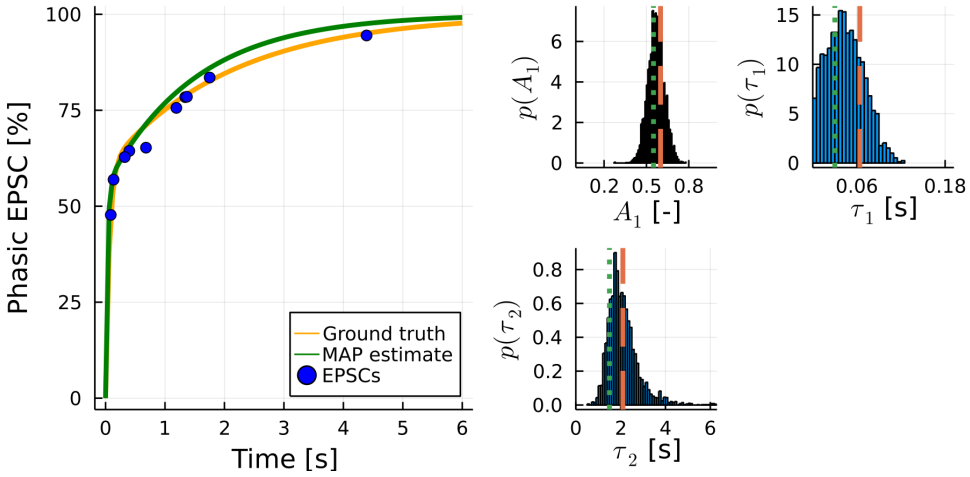


Figure 8.2: **ESB-BAL for optimal vesicle sampling.** ESB-BAL is being applied to a synthetic presynaptic terminal with ground-truth parameters $\tau_D^{(1)^*} = 64ms$, $\tau_D^{(2)^*} = 2.1s$, and $A_1{}^* = 0.6$ (i.e. same values as in [25]). Left panel: the orange curve corresponds to the ground-truth proportion of refilled vesicles as a function of time after presynaptic depletion $x(t)$. Parameters $\tau_D^{(1)}$, $\tau_D^{(2)}$, and A_1 can be estimated by fitting Eq. 8.1 on EPSCs obtained at different time intervals Δ_t (blue dots). Right panels: marginal posterior distribution of the parameters after 10 observations obtained via ESB-BAL. Distributions are updated after each new observation using the filter described in Section 5.4. Orange dashed lines: ground-truth values. Green dashed lines: MAP estimates. The green curve in the left panel corresponds to the estimate of $x(t)$ computed using the MAP estimates of $\tau_D^{(1)}$, $\tau_D^{(2)}$, and A_1.

8.4.3 Extensions to different experimental settings

In Chapter 5, we used Optimal Experiment Design to optimize the informativeness of presynaptic stimulation times. Our main assumption was that evoked spike times were the only experimental variables that could be modified by the experimentalists. Future work should consider the possibility to expand the space of modifiable experimental parameters, and to include not only the stimulation times, but also pharmacological means. For instance, in Chapter 6, we propose to also optimize the external calcium concentration of the experimental solution (which modifies the release probability of vesicles) to optimally observe either STD or quantal release.

[2] `https://github.com/camillegontier/OptimalVesicleSampling.jl`

Another experimental setting to which OED could be applied is Long-Term Potentiation. Especially, our proposed framework could be used to optimize the experimental design (namely the number of repetitions, the frequency of pre- and postsynaptic stimulations, or the time intervals Δ_t) to study STDP. Studying long-term synaptic plasticity involves comparing several candidate models (pair of spikes, triplet, or quadruplet rules; All-to-all or Nearest-Neighbor interactions; possible dependence of the weight change amplitude on the current weights; etc [198]) and to infer several parameters defining the shape of the STDP windows. This inference process could be optimized using OED.

8.4.4 Extensions to different fields

The overall goal of this doctoral thesis was to propose statistical tools and approaches to answer practical questions about synaptic characterization, i.e. how to quantify the discriminability between competing models (Chapter 4), how to compute the optimal presynaptic stimulation times (Chapter 5), how to pick the best experimental design for model selection (Chapter 6), and how to explain the organization of synaptic proteins from a normative perspective (Chapter 7). However, we argue that the solutions and approaches we proposed (e.g. criterion for identifiability, model fitting and parameter inference methods, OED method for model selection, etc.) are actually general and multi-purposed statistical tools, and could be well applied to different fields. For instance, our publication on practical identifiability has been mentioned in a recent study on synaptic parameters inference [222], but also in a paper studying models of how fake news propagate on social networks [223]. Similarly, the issue of model overfitting (and of the associated low accuracy of estimated parameters) has been mentioned in a recent paper on synaptic Short-Term Plasticity [224], but could be well applied to any other field using statistical modelling.

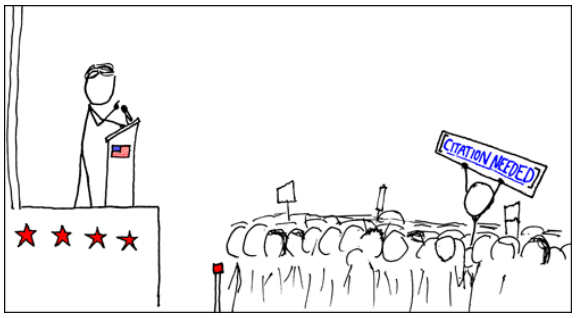


Figure 8.3: On correct bibliography management[3].

[3] `https://xkcd.com/285/`, use with authorization.

Bibliography

[1] Cary Soares, Daniel Trotter, André Longtin, Jean-Claude Béïque, and Richard Naud. Parsing out the variability of transmission at central synapses using optical quantal analysis. *Frontiers in synaptic neuroscience*, 11:22, 2019. v, 9, 16, 80, 93, 96

[2] Peter Dayan and Laurence F Abbott. Theoretical neuroscience: computational and mathematical modeling of neural systems. 2001. vi, 5, 8, 20

[3] Ola Bykowska, Camille Gontier, Anne-Lene Sax, David W Jia, Milton Llera Montero, Alex D Bird, Conor Houghton, Jean-Pascal Pfister, and Rui Ponte Costa. Model-based inference of synaptic transmission. *Frontiers in synaptic neuroscience*, 11:21, 2019. vi, 28, 42, 73, 117

[4] Camille Gontier, Jakob Jordan, and Mihai A Petrovici. Delaunay: a dataset of abstract art for psychophysical and machine learning research. *arXiv preprint arXiv:2201.12123*, 2022. vi

[5] Alex D Bird, Mark J Wall, and Magnus JE Richardson. Bayesian inference of synaptic quantal parameters from correlated vesicle release. *Frontiers in computational neuroscience*, 10:116, 2016. ix, 1, 6, 13, 16, 41, 43, 45, 46, 55, 56, 70, 74, 76, 77, 80, 81, 83, 154

[6] A Reyes and Bert Sakmann. Developmental switch in the short-term modification of unitary EPSPs evoked in layer 2/3 and layer 5 pyramidal neurons of rat neocortex. *THe Journal of Neuroscience*, 19(10):3827–3835, 1999. ix, 2, 119

[7] Alex Reyes, Rafael Lujan, Andrej Rozov, Nail Burnashev, Peter Somogyi, and Bert Sakmann. Target-cell-specific facilitation and depression in neocortical circuits. *Nature neuroscience*, 1(4):279–285, 1998. ix, 2, 119

[8] Dirk Feldmeyer, John F Wesseling, and P Jesper Sjöström. Methods for synaptic interrogation. *Frontiers in Synaptic Neuroscience*, 12:23, 2020. ix, 11

[9] J Del Castillo and B3 Katz. Quantal components of the end-plate potential. *The Journal of physiology*, 124(3):560–573, 1954. 3, 42, 56, 57, 71, 93, 96

[10] Wikipedia contributors. Synapse — Wikipedia, the free encyclopedia, 2022. URL `https://en.wikipedia.org/w/index.php?title=Synapse&oldid=1087147987`. [Online; accessed 10-June-2022]. 3

[11] Anu G Nair, Paola Muttathukunnel, and Martin Müller. Distinct molecular pathways govern presynaptic homeostatic plasticity. *Cell reports*, 37(11): 110105, 2021. 4, 100

[12] Bernard Katz. The release of neural transmitter substances. *Liverpool University Press*, pages 5–39, 1969. 3, 42, 74, 143

[13] Melissa Silva, Van Tran, and Alain Marty. Calcium-dependent docking of synaptic vesicles. *Trends in Neurosciences*, 44(7):579–592, 2021. 5

[14] Janet Barroso-Flores, Marco A Herrera-Valdez, Elvira Galarraga, and José Bargas. Models of short-term synaptic plasticity. *The Plastic Brain*, pages 41–57, 2017. 5, 8

[15] Matthias H Hennig. Theoretical models of synaptic short term plasticity. *Frontiers in computational neuroscience*, 7:45, 2013. 5, 6, 7, 10

[16] AW Liley and KAK North. An electrical investigation of effects of repetitive stimulation on mammalian neuromuscular junction. *Journal of neurophysiology*, 16(5):509–527, 1953. 5

[17] Misha Tsodyks, Klaus Pawelzik, and Henry Markram. Neural networks with dynamic synapses. *Neural computation*, 10(4):821–835, 1998. 6, 43, 74, 82, 143

[18] Jean-Pascal Pfister, Peter Dayan, and Máté Lengyel. Synapses with short-term plasticity are optimal estimators of presynaptic membrane potentials. *Nature neuroscience*, 13(10):1271–1275, 2010. 6

[19] Gianluigi Mongillo, Omri Barak, and Misha Tsodyks. Synaptic theory of working memory. *Science*, 319(5869):1543–1546, 2008. 6

[20] Alessandro Barri, Yun Wang, David Hansel, and Gianluigi Mongillo. Quantifying repetitive transmission at chemical synapses: a generative-model approach. *Eneuro*, 3(2), 2016. 6, 13, 16, 41, 43, 44, 45, 53, 55, 56, 57, 58, 70, 74, 76, 80, 82, 117, 149, 153

[21] Galit Fuhrmann, Anna Cowan, Idan Segev, Misha Tsodyks, and Christian Stricker. Multiple mechanisms govern the dynamics of depression at neocortical synapses of young rats. *The Journal of physiology*, 557(2):415–438, 2004. 7

[22] Shakuntala Savanthrapadian, Thomas Meyer, Claudio Elgueta, Sam A Booker, Imre Vida, and Marlene Bartos. Synaptic properties of som-and cck-expressing cells in dentate gyrus interneuron networks. *Journal of Neuroscience*, 34(24): 8197–8209, 2014. 7

[23] Jeremy S Dittman, Anatol C Kreitzer, and Wade G Regehr. Interplay between facilitation, depression, and residual calcium at three presynaptic terminals. *Journal of Neuroscience*, 20(4):1374–1385, 2000. 7

[24] Julia Trommershäuser, Ralf Schneggenburger, Annette Zippelius, and Erwin Neher. Heterogeneous Presynaptic Release Probabilities: Functional Relevance for Short-Term Plasticity. *Biophysical journal*, 84(3):1563–1579, March 2003. 9, 119

[25] Stefan Hallermann, Anna Fejtova, Hartmut Schmidt, Annika Weyhersmüller, R Angus Silver, Eckart D Gundelfinger, and Jens Eilers. Bassoon speeds vesicle reloading at a central excitatory synapse. *Neuron*, 68(4):710–723, 2010. 10, 119, 120

[26] Abed Ghanbari, Aleksey Malyshev, Maxim Volgushev, and Ian H Stevenson. Estimating short-term synaptic plasticity from pre-and postsynaptic spiking. *PLoS computational biology*, 13(9):e1005738, 2017. 10, 42, 56, 117

[27] B. Gutkin and F. Zeldenrust. Spike frequency adaptation. *Scholarpedia*, 9(2): 30643, 2014. doi: 10.4249/scholarpedia.30643. revision #143322. 10

[28] Takayuki Yamashita, Taro Ishikawa, and Tomoyuki Takahashi. Developmental increase in vesicular glutamate content does not cause saturation of ampa receptors at the calyx of held synapse. *Journal of Neuroscience*, 23(9):3633–3638, 2003. 10

[29] Jeremy Lewi, Robert Butera, and Liam Paninski. Sequential optimal design of neurophysiology experiments. *Neural computation*, 21(3):619–687, 2009. 10, 23, 70, 71, 72, 73, 117

[30] Madineh Sedigh-Sarvestani, Larry A Palmer, and Diego Contreras. Thalamocortical synapses in the cat visual system in vivo are weak and unreliable. *eLife*, 8, April 2019. 11, 36, 119

[31] William B Levy and Robert A Baxter. Energy-efficient neuronal computation via quantal synaptic failures. *Journal of Neuroscience*, 22(11):4746–4755, 2002. 11, 42, 80

[32] Graeme W Davis and Martin Müller. Homeostatic control of presynaptic neurotransmitter release. *Annual review of physiology*, 77:251–270, 2015. 11, 42, 61, 80

[33] Corinna Wentzel, Igor Delvendahl, Sebastian Sydlik, Oleg Georgiev, and Martin Müller. Dysbindin links presynaptic proteasome function to homeostatic recruitment of low release probability vesicles. *Nature communications*, 9(1): 1–16, 2018. 11, 42, 80, 100

[34] Rui Ponte Costa, Robert C Froemke, P Jesper Sjöström, and Mark CW van Rossum. Correction: Unified pre-and postsynaptic long-term plasticity enables reliable and flexible learning. *Elife*, 4:e11988, 2015. 11, 80

[35] Rui Ponte Costa, Beatriz EP Mizusaki, P Jesper Sjöström, and Mark CW van Rossum. Functional consequences of pre-and postsynaptic expression of synaptic plasticity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1715):20160153, 2017. 11, 42, 80

[36] Rui Ponte Costa, Zahid Padamsey, James A D'Amour, Nigel J Emptage, Robert C Froemke, and Tim P Vogels. Synaptic transmission optimization predicts expression loci of long-term plasticity. *Neuron*, 96(1):177–189, 2017. 11, 42, 80

[37] Myrrhe Van Spronsen and Casper C Hoogenraad. Synapse pathology in psychiatric and neurologic disease. *Current neurology and neuroscience reports*, 10 (3):207–214, 2010. 11, 42

[38] Robert D Terry, Eliezer Masliah, David P Salmon, Nelson Butters, Richard DeTeresa, Robert Hill, Lawrence A Hansen, and Robert Katzman. Physical basis of cognitive alterations in alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Annals of Neurology: Official Journal of*

the American Neurological Association and the Child Neurology Society, 30(4): 572–580, 1991. 11

[39] Milton Llera-Montero, João Sacramento, and Rui Ponte Costa. Computational roles of plastic probabilistic synapses. *Current Opinion in Neurobiology*, 54: 90–97, February 2019. 12, 29, 38, 119

[40] R Angus Silver. Estimation of nonuniform quantal parameters with multiple-probability fluctuation analysis: theory, application and limitations. *Journal of neuroscience methods*, 130(2):127–141, 2003. 12

[41] John D Clements. Variance–mean analysis: a simple and reliable approach for investigating synaptic transmission and modulation. *Journal of neuroscience methods*, 130(2):115–125, 2003. 12

[42] Volker Scheuss and Erwin Neher. Estimating synaptic parameters from mean, variance, and covariance in trains of synaptic responses. *Biophysical journal*, 81(4):1970–1989, 2001. 12, 74

[43] Alexander C Meyer, Erwin Neher, and Ralf Schneggenburger. Estimation of quantal size and number of functional active zones at the calyx of held synapse by nonstationary epsc variance analysis. *Journal of Neuroscience*, 21(20): 7889–7900, 2001. 12

[44] Ralf Schneggenburger, Alexander C Meyer, and Erwin Neher. Released fraction and total size of a pool of immediately available transmitter quanta at a calyx synapse. *Neuron*, 23(2):399–409, 1999. 12

[45] Monica S Thanawala and Wade G Regehr. Determining Synaptic Parameters Using High-Frequency Activation. *Journal of neuroscience methods*, March 2016. 12

[46] Alexander E Dityatev, Radik Sh Altinbaev, Andrej V Astrelin, and Leon L Voronin. Combining principal component and spectral analyses with the method of moments in studies of quantal transmission. *Journal of neuroscience methods*, 130(2):173–199, 2003. 12

[47] Christian Stricker and Stephen J Redman. Quantal analysis based on density estimation. *Journal of neuroscience methods*, 130(2):159–171, 2003. 13, 74

[48] Natalí Barros-Zulaica, John Rahmon, Giuseppe Chindemi, Rodrigo Perin, Henry Markram, Eilif Muller, and Srikanth Ramaswamy. Estimating the readily-releasable vesicle pool size at synaptic connections in the neocortex. *Frontiers in synaptic neuroscience*, page 29, 2019. 13

[49] Anna Kutschireiter, Simone Carlo Surace, and Jean-Pascal Pfister. The hitchhiker's guide to nonlinear filtering. *Journal of Mathematical Psychology*, 94: 102307, 2020. 13, 73

[50] D Anderson and K Burnham. Model selection and multi-model inference. *Second. NY: Springer-Verlag*, 63(2020):10, 2004. 14, 15

[51] William Bialek. Perspectives on theory at the interface of physics and biology. *arXiv preprint arXiv:1512.08954*, 2015. 14

[52] GEP Box. Robustness in the strategy of scientific model building, robustness in statistics, rl launer and gn wilkinson, eds. *Academic Press. Retrieved April*, 12:2017, 1979. 14

[53] Arturo Rosenblueth and Norbert Wiener. The role of models in science. *Philosophy of science*, 12(4):316–321, 1945. 14

[54] Christian Keysers, Valeria Gazzola, and Eric-Jan Wagenmakers. Using bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature neuroscience*, 23(7):788–799, 2020. 15

[55] Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978. 15, 51, 147

[56] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974. 15

[57] Kenneth P Burnham and David R Anderson. Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, 33(2): 261–304, 2004. 15, 59

[58] Sumio Watanabe and Manfred Opper. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, 11(12), 2010. 16

[59] Gerda Claeskens and Nils Lid Hjort. The focused information criterion. *Journal of the American Statistical Association*, 98(464):900–916, 2003. 16

[60] Edward J Hannan and Barry G Quinn. The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):190–195, 1979. 16

[61] Gilles Celeux and Jean-Baptiste Durand. Selecting hidden markov model state number with cross-validated likelihood. *Computational Statistics*, 23(4): 541–564, 2008. 16

[62] Mervyn Stone. An asymptotic equivalence of choice of model by cross-validation and akaike's criterion. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):44–47, 1977. 16

[63] Jun Shao. An asymptotic theory for linear model selection. *Statistica sinica*, pages 221–242, 1997. 16

[64] Klaas Enno Stephan, Will D Penny, Jean Daunizeau, Rosalyn J Moran, and Karl J Friston. Bayesian model selection for group studies. *Neuroimage*, 46(4): 1004–1017, 2009. 16

[65] Alan F Heavens, TD Kitching, and L Verde. On model selection forecasting, dark energy and modified gravity. *Monthly Notices of the Royal Astronomical Society*, 380(3):1029–1035, 2007. 16, 117, 152

[66] Ben Placek, Kevin H Knuth, and Daniel Angerhausen. Exonest: Bayesian model selection applied to the detection and characterization of exoplanets via photometric variations. *The Astrophysical Journal*, 795(2):112, 2014. 16

[67] Tiago M Fragoso, Wesley Bertoli, and Francisco Louzada. Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review*, 86(1):1–28, 2018. 16

[68] David JC MacKay and David JC Mac Kay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003. 17, 49

[69] P Fatt and B Katz. Spontaneous subthreshold activity at motor nerve endings. *The Journal of physiology*, 117(1):109–128, 1952. 18, 57

[70] Andreas Raue, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009. 19, 41, 47, 48, 49, 50, 59, 117

[71] Michał Komorowski, Maria J Costa, David A Rand, and Michael PH Stumpf. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *Proceedings of the National Academy of Sciences*, 108(21):8645–8650, 2011. 19, 20, 41

[72] A Raue, C Kreutz, T Maiwald, U Klingmuller, and J Timmer. Addressing parameter identifiability by model-based experimentation. *IET systems biology*, 5(2):120–130, 2011. 19, 41, 117

[73] Shinsuke Koyama. Coding efficiency and detectability of rate fluctuations with non-poisson neuronal firing. In *Advances in Neural Information Processing Systems*, pages 180–188, 2012. 19, 41, 59

[74] Keegan E Hines, Thomas R Middendorf, and Richard W Aldrich. Determination of parameter identifiability in nonlinear biophysical models: A bayesian approach. *The Journal of general physiology*, 143(3):401–416, 2014. 19, 41

[75] Gemma Massonis and Alejandro F Villaverde. Finding and breaking lie symmetries: Implications for structural identifiability and observability in biological modelling. *Symmetry*, 12(3):469, 2020. 19, 41, 47

[76] DJ Venzon and SH Moolgavkar. A method for computing profile-likelihood-based confidence intervals. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 37(1):87–94, 1988. 19, 48

[77] Susan A Murphy and Aad W Van der Vaart. On profile likelihood. *Journal of the American Statistical Association*, 95(450):449–465, 2000. 19, 48

[78] Dennis V Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, pages 986–1005, 1956. 22, 71

[79] Xun Huan and Youssef M Marzouk. Simulation-based optimal bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232 (1):288–317, 2013. 22, 23, 61, 71, 73

[80] Paola Sebastiani and Henry P Wynn. Maximum entropy sampling and optimal bayesian experimental design. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):145–157, 2000. 23, 70

[81] Adam Foster, Martin Jankowiak, Eli Bingham, Paul Horsfall, Yee Whye Teh, Tom Rainforth, and Noah Goodman. Variational bayesian optimal experimental design. *arXiv preprint arXiv:1903.05480*, 2019. 23, 71, 73, 117

[82] Joep Vanlier, Christian A Tiemann, Peter AJ Hilbers, and Natal AW van Riel. Optimal experiment design for model selection in biochemical networks. *BMC systems biology*, 8(1):20, 2014. 23, 61, 88, 89

[83] Stefano Balietti, Brennan Klein, and Christoph Riedl. Fast model-selection through adapting design of experiments maximizing information gain. *arXiv preprint arXiv:1807.07024*, 2018. 23, 61, 89

[84] S Nabavi, R Fox, C D Proulx, J Y Lin, and R Y Tsien. Engineering a memory with LTD and LTP. *Nature*, 2014. 29

[85] Rui Ponte Costa, Zahid Padamsey, James A D'amour, Nigel J Emptage, Robert C Froemke, and Tim P Vogels. Synaptic Transmission Optimization Predicts Expression Loci of Long-Term Plasticity. *Neuron*, 96(1):177–189.e7, September 2017. 29, 38

[86] Pieter R Roelfsema and Anthony Holtmaat. Control of synaptic plasticity in deep cortical networks. *Nature Reviews Neuroscience*, 19(3):166–180, March 2018. 29

[87] Leena E Williams and Anthony Holtmaat. Higher-Order Thalamocortical Inputs Gate Synaptic Long-Term Potentiation via Disinhibition. *Neuron*, 0(0), November 2018. 29

[88] Gerardo Malagon, Takafumi Miki, Isabel Llano, Erwin Neher, and Alain Marty. Counting Vesicular Release Events Reveals Binomial Release Statistics at Single Glutamatergic Synapses. *The Journal of neuroscience*, 36(14):4010–4025, April 2016. 29

[89] Henry Markram, Y Wang, and Misha Tsodyks. Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Natl. Acad. Sci. USA*, 95 (9):5323–5328, 1998. 29, 33, 34

[90] Robert S Zucker and Wade G Regehr. Short-term synaptic plasticity. *Annu Rev Physiol*, 64(1):355–405, January 2002. 29

[91] D S Faber and H Korn. Applicability of the coefficient of variation method for analyzing synaptic plasticity. *Biophysical journal*, 60(5):1288–1294, November 1991. 29

[92] S F Traynelis, R A Silver, and S G Cull-Candy. Estimated conductance of glutamate receptor channels activated during EPSCs at the cerebellar mossy fiber-granule cell synapse. *Neuron*, 11(2):279–289, August 1993. 29

[93] Misha Tsodyks and Henry Markram. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl. Acad. Sci. USA*, 94(2):719–723, 1997. 29, 34

[94] Lynn E Dobrunz and Charles F Stevens. Response of hippocampal synapses to natural stimulation patterns. *Neuron*, 22(1):157–166, 1999. 29

[95] J T R Isaac, Katherine A Buchanan, R U Muller, and J R Mellor. Hippocampal Place Cell Firing Patterns Can Induce Long-Term Synaptic Plasticity In Vitro. *The Journal of neuroscience*, 29(21):6840–6850, May 2009. 29

[96] J Del Castillo and B Katz. Quantal components of the end-plate potential. *The Journal of physiology*, 124(3):560–573, 1954. 29

[97] Alan Larkman, Ken Stratford, and Julian Jack. Quantal analysis of excitatory synaptic action and depression in hippocampal slices. *Nature*, 350(6316): 344–347, March 1991. 29

[98] Frederic Lanore and R Angus Silver. Extracting Quantal Properties of Transmission at Central Synapses. In *Advanced Patch-Clamp Analysis for Neuroscientists*, pages 193–211. Springer New York, New York, NY, January 2016. 29, 30, 33, 38

[99] Henry Markram, J Lübke, M Frotscher, A Roth, and B Sakmann. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *The Journal of physiology*, 500 (Pt 2):409–440, April 1997. 30

[100] Rui Ponte Costa, Robert C Froemke, P Jesper Sjostrom, and Mark C W van Rossum. Unified pre- and postsynaptic long-term plasticity enables reliable and flexible learning. *eLife*, 4:e09457, 2015. 30, 36, 38

[101] Gardave Singh Bhumbra and Marco Beato. Reliable evaluation of the quantal determinants of synaptic efficacy using Bayesian analysis. *Journal of Neurophysiology*, 109(2):603–620, 2013. 30, 33, 37

[102] Rui P Costa, P Jesper Sjostrom, and Mark C W van Rossum. Probabilistic inference of short-term synaptic plasticity in neocortical microcircuits. *Frontiers in Computational Neuroscience*, 7:75, 2013. 32, 33, 34, 35, 37, 38

[103] JV Le Bé and Henry Markram. Spontaneous and evoked synaptic rewiring in the neonatal neocortex. *Proceedings of the National Academy of Sciences*, 103 (35):13214, 2006. 33

[104] Henry Markram. The blue brain project. *Nature Reviews Neuroscience*, 7(2): 153–160, February 2006. 33

[105] Yun Wang, Henry Markram, Philip H Goodman, Thomas K Berger, Junying Ma, and Patricia S Goldman-Rakic. Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nature Publishing Group*, 9(4):534–542, April 2006. 33

[106] Tania Rinaldi, Gilad Silberberg, and Henry Markram. Hyperconnectivity of local neocortical microcircuitry induced by prenatal exposure to valproic acid. *Cerebral Cortex*, 18(4):763–770, April 2008. 33

[107] Srikanth Ramaswamy, Sean L Hill, James G King, Felix Schürmann, Yun Wang, and Henry Markram. Intrinsic morphological diversity of thick-tufted layer 5 pyramidal neurons ensures robust and invariant properties of in silico synaptic connections. *The Journal of physiology*, pages no–no, January 2012. 33

[108] G Testa-Silva, A Loebel, M Giugliano, C P J de Kock, H D Mansvelder, and R M Meredith. Hyperconnectivity and Slow Synapses during Early Development of Medial Prefrontal Cortex in a Mouse Model for Mental Retardation and Autism. *Cerebral Cortex*, 22(6):1333–1342, May 2012. 33

[109] Armando Romani, Cristina Marchetti, Daniela Bianchi, Xavier Leinekugel, Panayiota Poirazi, Michele Migliore, and Hélène Marie. Computational modeling of the effects of amyloid-beta on release probability at hippocampal synapses. *Frontiers in Computational Neuroscience*, 7, 2013. 33

[110] Alex Loebel, Gilad Silberberg, Daniela Helbig, Henry Markram, Misha Tsodyks, and Magnus J E Richardson. Multiquantal release underlies the distribution of synaptic efficacies in the neocortex. *Frontiers in Computational Neuroscience*, 3:27, 2009. 33, 34, 36

[111] A Barri, Y Wang, D Hansel, and G Mongillo. Quantifying Repetitive Transmission at Chemical Synapses: A Generative-Model Approach. *eneuro*, 3(2), March 2016. 33, 34, 35, 36

[112] Alex D Bird, Mark J Wall, and Magnus J E Richardson. Bayesian Inference of Synaptic Quantal Parameters from Correlated Vesicle Release. *Frontiers in Computational Neuroscience*, 10:116, 2016. 33, 34, 35, 36, 37, 38

[113] Abed Ghanbari, Aleksey Malyshev, Maxim Volgushev, and Ian H Stevenson. Estimating short-term synaptic plasticity from pre- and postsynaptic spiking. *PLOS Computational Biology*, 13(9):e1005738, September 2017. 33, 35, 37, 38

[114] Matthias H Hennig. Theoretical models of synaptic short term plasticity. *Frontiers in Computational Neuroscience*, pages 1–10, April 2013. 33

[115] Juan Varela, Kamal Sen, Jay Gibson, Joshua Fost, L F Abbott, and Sacha Nelson. A Quantitative Description of Short-Term Plasticity at Excitatory Synapses in Layer 2/3 of Rat Primary Visual Cortex. *The Journal of neuroscience*, 17(20):7926, October 1997. 33

[116] Simen Tennøe, Geir Halnes, and Gaute T Einevoll. Uncertainpy: A Python Toolbox for Uncertainty Quantification and Sensitivity Analysis in Computational Neuroscience. *Frontiers in neuroinformatics*, 12:e94, August 2018. 33

[117] Emily Stone, Heikki Haario, and J. Josh Lawrence. A kinetic model for the frequency dependence of cholinergic modulation at hippocampal GABAergic synapses. *Mathematical Biosciences*, 258:162–175, 2014. ISSN 18793134. doi: 10.1016/j.mbs.2014.09.013. URL `http://dx.doi.org/10.1016/j.mbs.2014.09.013`. 34

[118] Aurélie Pala and Carl C H Petersen. In Vivo Measurement of Cell-Type-Specific Synaptic Connectivity and Synaptic Transmission in Layer 2/3 Mouse Barrel Cortex. *Neuron*, 85(1):68–75, January 2015. 36

[119] Paolo Puggioni, Marta Jelitai, Ian Duguid, and Mark C W van Rossum. Extraction of Synaptic Input Properties in Vivo. *Neural Computation*, 29(7): 1745–1768, June 2017. 36

[120] Kenneth Latimer, Fred Rieke, and Jonathan W Pillow. Inferring synaptic inputs from spikes with a conductance-based neural encoding model. *bioRxiv*, page 281089, March 2018. 36

[121] R Brette and Wulfram Gerstner. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology*, 94 (5):3637, 2005. 37

[122] A Ghanbari, N Ren, C Keine, C Stoelzel, B Englitz bioRxiv, and 2018. Functional connectivity with short-term dynamics explains diverse patterns of excitatory spike transmission in vivo. *biorxiv.org*, 2018. 37

[123] Therese Abrahamsson, Christina You Chien Chou, Si Ying Li, Adamo Mancino, Rui Ponte Costa, Jennifer Anne Brock, Erin Nuro, Katherine Anne Buchanan, Dale Elgar, Arne Vladimir Blackman, and others. Differential regulation of evoked and spontaneous release by presynaptic NMDA receptors. *Neuron*, 96 (4):839–855, 2017. 37

[124] N Barros-Zulaica, J Rahmon, G Chindemi, R Perin, Henry Markram, S Ramaswamy, and E Muller. Estimating the Readily-Releasable Vesicle Pool Size at Synaptic Connections in a Neocortical Microcircuit. *bioRxiv*, 215:646497, May 2019. 37

[125] B Walmsley, F R Edwards, and D J Tracey. Nonuniform release probabilities underlie quantal synaptic transmission at a mammalian excitatory central synapse. *Journal of Neurophysiology*, 60(3):889–908, September 1988. 38

[126] Stephanie A Rey, Catherine A Smith, Milena W Fowler, Freya Crawford, Jemima J Burden, and Kevin Staras. Ultrastructural and functional fate of recycled vesicles in hippocampal synapses. *Nature Communications*, 6(1):8043, August 2015. 38

[127] Ai-Hui Tang, Haiwen Chen, Tuo P Li, Sarah R Metzbower, Harold D MacGillavry, and Thomas A Blanpied. A trans-synaptic nanocolumn aligns neurotransmitter release to receptors. *Nature*, 536(7615):210–214, August 2016. 38

[128] Nordine Helassa, Céline D Dürst, Catherine Coates, Silke Kerruth, Urwa Arif, Christian Schulze, J Simon Wiegert, Michael Geeves, Thomas G Oertner, and Katalin Torok. Ultrafast glutamate sensors resolve high-frequency release at Schaffer collateral synapses. *Proceedings of the National Academy of Sciences of the United States of America*, May 2018. 38

[129] Cary Soares, Daniel Trotter, Andre Longtin, Jean-Claude Béïque, and Richard Naud. Parsing out the variability of transmission at central synapses using optical quantal analysis. *bioRxiv*, page 624692, May 2019. 38

[130] Kiryl D Piatkevich, Seth Bensussen, Hua-an Tseng, Sanaya N Shroff, Violeta Gisselle Lopez-Huerta, Demian Park, Erica E Jung, Or A Shemesh, Christoph Straub, Howard J Gritton, Michael F Romano, Emma Costa, Bernardo L Sabatini, Zhanyan Fu, Edward S Boyden, and Xue Han. Population imaging of neural activity in awake behaving mice in multiple brain regions. *bioRxiv*, 350:616094, April 2019. 38

[131] James J Jun, Nicholas A Steinmetz, Joshua H Siegle, Daniel J Denman, Marius Bauza, Brian Barbarits, Albert K Lee, Costas A Anastassiou, Alexandru Andrei, Çağatay Aydın, Mladen Barbic, Timothy J Blanche, Vincent Bonin, João Couto, Barundeb Dutta, Sergey L Gratiy, Diego A Gutnisky, Michael Hausser,

Bill Karsh, Peter Ledochowitsch, Carolina Mora Lopez, Catalin Mitelut, Silke Musa, Michael Okun, Marius Pachitariu, Jan Putzeys, P Dylan Rich, Cyrille Rossant, Wei-lung Sun, Karel Svoboda, Matteo Carandini, Kenneth D Harris, Christof Koch, John O'Keefe, and Timothy D Harris. Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232–236, November 2017. 38

[132] Jean-Pascal Pfister, Peter Dayan, and Máté Lengyel. Synapses with short-term plasticity are optimal estimators of presynaptic membrane potentials. *Nature Neuroscience*, 13(10):1271–1275, October 2010. 38

[133] Johanna S Jackson, Jonathan Witton, James D Johnson, Zeshan Ahmed, Mark Ward, Andrew D Randall, Michael L Hutton, John T Isaac, Michael J O'Neill, and Michael C Ashby. Altered Synapse Stability in the Early Stages of Tauopathy. *Cell reports*, 18(13):3063–3068, March 2017. 38

[134] Camille Gontier and Jean-Pascal Pfister. Identifiability of a binomial synapse. *Frontiers in computational neuroscience*, 14:86, 2020. 39, 71, 74, 80, 81, 117, 143

[135] Nathaniel D Daw et al. Trial-by-trial data analysis using computational models. *Decision making, affect, and learning: Attention and performance XXIII*, 23 (1), 2011. 40, 52

[136] Abed Ghanbari, Naixin Ren, Christian Keine, Carl Stoelzel, Bernhard Englitz, Harvey Swadlow, and Ian Stevenson. Functional connectivity with short-term dynamics explains diverse patterns of excitatory spike transmission in vivo. *bioRxiv*, page 475178, 2018. 42

[137] Rui Ponte Costa, Robert C Froemke, P Jesper Sjöström, and Mark CW van Rossum. Unified pre-and postsynaptic long-term plasticity enables reliable and flexible learning. *Elife*, 4:e09457, 2015. 42, 43, 61

[138] Daqing Guo and Chunguang Li. Stochastic resonance in hodgkin–huxley neuron induced by unreliable synaptic transmission. *Journal of theoretical biology*, 308: 105–114, 2012. 42, 80

[139] Brad E Pfeiffer and Kimberly M Huber. The state of synapses in fragile x syndrome. *The Neuroscientist*, 15(5):549–567, 2009. 42

[140] Klaas E Stephan, Torsten Baldeweg, and Karl J Friston. Synaptic plasticity and dysconnection in schizophrenia. *Biological psychiatry*, 59(10):929–939, 2006. 42

[141] Paolo Calabresi, Barbara Picconi, Lucilla Parnetti, and Massimiliano Di Filippo. A convergent model for cognitive dysfunctions in parkinson's disease: the critical dopamine–acetylcholine synaptic balance. *The Lancet Neurology*, 5(11): 974–983, 2006. 42

[142] Thomas C Südhof. Neuroligins and neurexins link synaptic function to cognitive disease. *Nature*, 455(7215):903–911, 2008. 42

[143] Dennis J Selkoe. Alzheimer's disease is a synaptic failure. *Science*, 298(5594): 789–791, 2002. 42

[144] Jeffrey M Welch, Jing Lu, Ramona M Rodriguiz, Nicholas C Trotta, Joao Peca, Jin-Dong Ding, Catia Feliciano, Meng Chen, J Paige Adams, Jianhong Luo, et al. Cortico-striatal synaptic defects and ocd-like behaviours in sapap3-mutant mice. *Nature*, 448(7156):894–900, 2007. 42

[145] Julie A Kauer and Robert C Malenka. Synaptic plasticity and addiction. *Nature reviews neuroscience*, 8(11):844–858, 2007. 42

[146] JM Bekkers, GB Richerson, and CF Stevens. Origin of variability in quantal size in cultured hippocampal neurons and hippocampal slices. *Proceedings of the National Academy of Sciences*, 87(14):5359–5362, 1990. 44

[147] Gardave Singh Bhumbra and Marco Beato. Reliable evaluation of the quantal determinants of synaptic efficacy using bayesian analysis. *Journal of neurophysiology*, 109(2):603–620, 2012. 44

[148] Michael C Edwards and Robert C MacCallum. *Current topics in the theory and application of latent variable models*. Routledge, 2012. 47, 59, 118

[149] John Mordechai Gottman. *The analysis of change*. Psychology Press, 1995. 47

[150] Henry Teicher. Identifiability of finite mixtures. *The annals of Mathematical statistics*, pages 1265–1269, 1963. 47

[151] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995. 49

[152] William H Jefferys and James O Berger. Sharpening ockham's razor on a bayesian strop. *Tech nical Report*, 1991. 49

[153] Daniel J Navarro, Mark A Pitt, and In Jae Myung. Assessing the distinguishability of models and the informativeness of data. *Cognitive psychology*, 49(1): 47–84, 2004. 51, 117

[154] Luigi Acerbi, Wei Ji Ma, and Sethu Vijayakumar. A framework for testing identifiability of bayesian models of perception. In *Advances in neural information processing systems*, pages 1026–1034, 2014. 51, 117

[155] Martin D Weinberg et al. Computing the bayes factor from a markov chain monte carlo simulation of the posterior distribution. *Bayesian Analysis*, 7(3): 737–770, 2012. 51

[156] Eric-Jan Wagenmakers, Tom Lodewyckx, Himanshu Kuriyal, and Raoul Grasman. Bayesian hypothesis testing for psychologists: A tutorial on the savage–dickey method. *Cognitive psychology*, 60(3):158–189, 2010. 51

[157] A Mootoovaloo, Bruce A Bassett, and M Kunz. Bayes factors via savage-dickey supermodels. *arXiv preprint arXiv:1609.02186*, 2016. 51

[158] James C Spall. Monte carlo computation of the fisher information matrix in nonstandard settings. *Journal of Computational and Graphical Statistics*, 14 (4):889–909, 2005. 52, 148

[159] Rui P Costa, P Jesper Sjostrom, and Mark CW Van Rossum. Probabilistic inference of short-term synaptic plasticity in neocortical microcircuits. *Frontiers in computational neuroscience*, 7:75, 2013. 55, 76, 79, 80, 117, 149

[160] Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207): 995–999, 2008. 59, 118

[161] Ryota Kobayashi, Yasuhiro Tsubo, and Shigeru Shinomoto. Made-to-order spiking neuron model equipped with a multi-timescale adaptive threshold. *Frontiers in computational neuroscience*, 3:9, 2009. 59, 118

[162] Alexandre Melanson, Jorge F Mejias, James J Jun, Leonard Maler, and André Longtin. A phenomenological model for self-initiated movement in electric fish. *BMC Neuroscience*, 15(S1):P112, 2014. 59, 118

[163] Tao Wang, Luping Yin, Xiaolong Zou, Yousheng Shu, Malte J Rasch, and Si Wu. A phenomenological synapse model for asynchronous neurotransmitter release. *Frontiers in computational neuroscience*, 9:153, 2016. 59, 118

[164] Daniel Levenstein, Veronica A Alvarez, Asohan Amarasingham, Habiba Azab, Richard C Gerkin, Andrea Hasenstaub, Ramakrishnan Iyer, Renaud B Jolivet, Sarah Marzen, Joseph D Monaco, et al. On the role of theory and modeling in neuroscience. *arXiv preprint arXiv:2003.13825*, 2020. 59, 118

[165] Harry L Van Trees. *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory.* John Wiley & Sons, 2004. 60

[166] Harry L Van Trees and Kristine L Bell. Bayesian bounds for parameter estimation and nonlinear filtering/tracking. *AMC*, 10:12, 2007. 60

[167] Lubomir Kostal, Petr Lansky, and Stevan Pilarski. Performance breakdown in optimal stimulus decoding. *Journal of Neural Engineering*, 12(3):036012, 2015. 60

[168] Gardave Singh Bhumbra and Marco Beato. Reliable evaluation of the quantal determinants of synaptic efficacy using bayesian analysis. *Journal of neurophysiology*, 109(2):603–620, 2013. 60, 80, 93, 94, 96, 97

[169] Christof Koch. *Biophysics of computation: information processing in single neurons.* Oxford university press, 2004. 61

[170] Renaud Jolivet, Ryota Kobayashi, Alexander Rauch, Richard Naud, Shigeru Shinomoto, and Wulfram Gerstner. A benchmark test for a quantitative assessment of simple neuron models. *Journal of neuroscience methods*, 169(2): 417–424, 2008. 61

[171] Wulfram Gerstner and Richard Naud. How good are neuron models? *Science*, 326(5951):379–380, 2009. 61

[172] Skander Mensi, Richard Naud, Christian Pozzorini, Michael Avermann, Carl CH Petersen, and Wulfram Gerstner. Parameter extraction and classification of three cortical neuron types reveals two distinct adaptation mechanisms. *Journal of neurophysiology*, 107(6):1756–1775, 2012. 61

[173] Alexandre René, André Longtin, and Jakob H Macke. Inference of a mesoscopic population model from population spike trains. *Neural Computation*, pages 1–51, 2020. 61

[174] Ralf Schneggenburger and Erwin Neher. Intracellular calcium dependence of transmitter release rates at a fast central synapse. *Nature*, 406(6798):889–893, 2000. 61

[175] Xuelin Lou, Volker Scheuss, and Ralf Schneggenburger. Allosteric modulation of the presynaptic ca 2+ sensor for vesicle fusion. *Nature*, 435(7041):497–501, 2005. 61

[176] Jianyuan Sun, Zhiping P Pang, Dengkui Qin, Abigail T Fahim, Roberto Adachi, and Thomas C Südhof. A dual-ca 2+-sensor model for neurotransmitter release in a central synapse. *Nature*, 450(7170):676–682, 2007. 61

[177] Guillaume Drion, Timothy O'Leary, and Eve Marder. Ion channel degeneracy enables robust and tunable neuronal firing rates. *Proceedings of the National Academy of Sciences*, 112(38):E5361–E5370, 2015. 61

[178] Rahul K Rathour and Rishikesh Narayanan. Degeneracy in hippocampal physiology and plasticity. *Hippocampus*, 29(10):980–1022, 2019. 61

[179] SP Asprey and S Macchietto. Statistical tools for optimal dynamic model building. *Computers & Chemical Engineering*, 24(2-7):1261–1267, 2000. 61

[180] Camille Gontier, Simone Carlo Surace, Igor Delvendahl, Martin Müller, and Jean-Pascal Pfister. Efficient sampling-based bayesian active learning, 2022. URL `https://arxiv.org/abs/2201.07539`. 69

[181] Timo Flesch, Jan Balaguer, Ronald Dekker, Hamed Nili, and Christopher Summerfield. Comparing continual task learning in minds and machines. *Proceedings of the National Academy of Sciences*, 115(44):E10313–E10322, 2018. 70

[182] Ashley F Emery and Aleksey V Nenarokomov. Optimal experiment design. *Measurement Science and Technology*, 9(6):864, 1998. 70

[183] Elizabeth G Ryan, Christopher C Drovandi, James M McGree, and Anthony N Pettitt. A review of modern computational algorithms for bayesian optimal design. *International Statistical Review*, 84(1):128–154, 2016. 70, 78

[184] Mijung Park and Jonathan Pillow. Bayesian active learning with localized priors for fast receptive field characterization. *Advances in neural information processing systems*, 25:2348–2356, 2012. 70, 71, 117

[185] Mijung Park, Greg Horwitz, and Jonathan W Pillow. Active learning of neural response functions with gaussian processes. In *NIPS*, pages 2043–2051. Citeseer, 2011. 70, 117

[186] Dan Crisan, Joaquin Miguez, et al. Nested particle filters for online parameter estimation in discrete-time state-space markov models. *Bernoulli*, 24(4A): 3039–3086, 2018. 71, 73, 81, 117, 144, 155

[187] Tim Besard, Christophe Foket, and Bjorn De Sutter. Effective extensible programming: Unleashing Julia on GPUs. *IEEE Transactions on Parallel and Distributed Systems*, 2018. ISSN 1045-9219. doi: 10.1109/TPDS.2018.2872064. 71, 74, 86

[188] Tim Besard, Valentin Churavy, Alan Edelman, and Bjorn De Sutter. Rapid software prototyping for heterogeneous and distributed platforms. *Advances in Engineering Software*, 132:29–46, 2019. 71, 74, 86

[189] Franz-Georg Wieland, Adrian L Hauber, Marcus Rosenblatt, Christian Tönsing, and Jens Timmer. On structural and practical identifiability. *Current Opinion in Systems Biology*, 2021. 71, 117

[190] Aditi Jha, Zoe C Ashwood, and Jonathan W Pillow. Bayesian active learning for discrete latent variable models. *arXiv preprint arXiv:2202.13426*, 2022. 73, 74, 117, 148

[191] Luigi Acerbi. Variational bayesian monte carlo. *Advances in Neural Information Processing Systems*, 31, 2018. 74

[192] Christopher C Drovandi, Minh-Ngoc Tran, et al. Improving the efficiency of fully bayesian optimal design of experiments using randomised quasi-monte carlo. *Bayesian Analysis*, 13(1):139–162, 2018. 78

[193] Henry Markram, Dimitri Pikus, Anirudh Gupta, and Misha Tsodyks. Potential for multiple mechanisms, phenomena and algorithms for synaptic plasticity at single synapses. *Neuropharmacology*, 37(4-5):489–500, 1998. 78

[194] William Stanley Jevons. The coal question. *An Inquiry Concerning the Prog*, 1862. 80

[195] Liam Paninski. Asymptotic theory of information-theoretic experimental design. *Neural Computation*, 17(7):1480–1507, 2005. 80, 92

[196] Rubén Moreno-Bote, Jeffrey Beck, Ingmar Kanitscheider, Xaq Pitkow, Peter Latham, and Alexandre Pouget. Information-limiting correlations. *Nature neuroscience*, 17(10):1410–1417, 2014. 80

[197] Laurence Aitchison, Jannes Jegminat, Jorge Aurelio Menendez, Jean-Pascal Pfister, Alexandre Pouget, and Peter E Latham. Synaptic plasticity as bayesian inference. *Nature Neuroscience*, 24(4):565–571, 2021. 80

[198] Jean-Pascal Pfister and Wulfram Gerstner. Triplets of spikes in a model of spike timing-dependent plasticity. *Journal of Neuroscience*, 26(38):9673–9682, 2006. 80, 121

[199] Luke Campagnola, Stephanie C Seeman, Thomas Chartrand, Lisa Kim, Alex Hoggarth, Clare Gamlin, Shinya Ito, Jessica Trinh, Pasha Davoudian, Cristina Radaelli, et al. Connectivity and synaptic physiology in the mouse and human neocortex. *bioRxiv*, 2021. 80

[200] Mahasen B Dehideniya, Christopher C Drovandi, and James M McGree. Optimal bayesian design for discriminating between models with intractable likelihoods in epidemiology. *Computational Statistics & Data Analysis*, 124:277–297, 2018. 88, 89

[201] Markus Hainy, David J Price, Olivier Restif, and Christopher Drovandi. Optimal bayesian design for model discrimination via classification. *Statistics and computing*, 32(2):1–17, 2022. 88, 89

[202] James M McGree. Developments of the total entropy utility function for the dual purpose of model discrimination and parameter estimation in bayesian design. *Computational Statistics & Data Analysis*, 113:207–225, 2017. 88, 89

[203] Mahasen B Dehideniya, Christopher C Drovandi, and James M McGree. Dual purpose bayesian design for parameter estimation and model discrimination in epidemiology using a synthetic likelihood approach. 2018. 89

[204] Anthony C Atkinson. Dt-optimum designs for model discrimination and parameter estimation. *Journal of Statistical planning and Inference*, 138(1):56–64, 2008. 89

[205] Yoshinori Sahara and Tomoyuki Takahashi. Quantal components of the excitatory postsynaptic currents at a rat central auditory synapse. *The Journal of physiology*, 536(1):189–197, 2001. 93, 96

[206] Ai-Hui Tang, Haiwen Chen, Tuo P Li, Sarah R Metzbower, Harold D MacGillavry, and Thomas A Blanpied. A trans-synaptic nanocolumn aligns neurotransmitter release to receptors. *Nature*, 536(7615):210–214, 2016. 99

[207] A Aldo Faisal, John A White, and Simon B Laughlin. Ion-channel noise places limits on the miniaturization of the brain's wiring. *Current Biology*, 15(12): 1143–1149, 2005. 99

[208] Mathias A Böhme, Christina Beis, Suneel Reddy-Alla, Eric Reynolds, Malou M Mampell, Andreas T Grasskamp, Janine Lützkendorf, Dominique Dufour Bergeron, Jan H Driller, Husam Babikir, et al. Active zone scaffolds differentially accumulate unc13 isoforms to tune ca2+ channel–vesicle coupling. *Nature neuroscience*, 19(10):1311–1320, 2016. 99

[209] Hirokazu Sakamoto, Tetsuroh Ariyoshi, Naoya Kimpara, Kohtaroh Sugao, Isamu Taiko, Kenji Takikawa, Daisuke Asanuma, Shigeyuki Namiki, and Kenzo Hirose. Synaptic weight set by munc13-1 supramolecular assemblies. *Nature neuroscience*, 21(1):41–49, 2018. 100

[210] Giuseppe Vicidomini, Andreas Schönle, Haisen Ta, Kyu Young Han, Gael Moneron, Christian Eggeling, and Stefan W Hell. Sted nanoscopy with time-gated detection: theoretical and experimental aspects. *PloS one*, 8(1):e54421, 2013. 100

[211] Adrian Baddeley and Rolf Turner. Spatstat: an r package for analyzing spatial point patterns. *Journal of statistical software*, 12:1–42, 2005. 101

[212] Adrian Baddeley et al. Analysing spatial point patterns in r. In *Workshop notes version*, volume 3, 2008. 101, 102

[213] Mark Berman and T Rolf Turner. Approximating point process likelihoods with glim. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41 (1):31–38, 1992. 102

[214] Deepak Nair, Eric Hosy, Jennifer D Petersen, Audrey Constals, Gregory Giannone, Daniel Choquet, and Jean-Baptiste Sibarita. Super-resolution imaging reveals that ampa receptors inside synapses are dynamically organized in nanodomains regulated by psd95. *Journal of Neuroscience*, 33(32):13204–13224, 2013. 104, 106, 107, 108

[215] Shuo Li, Sumana Raychaudhuri, Stephen Alexander Lee, Marisa M Brockmann, Jing Wang, Grant Kusick, Christine Prater, Sarah Syed, Hanieh Falahati, Raul Ramos, et al. Asynchronous release sites align with nmda receptors in mouse hippocampal synapses. *Nature communications*, 12(1):1–13, 2021. 104, 105

[216] Sridhar Raghavachari and John E Lisman. Properties of quantal transmission at ca1 synapses. *Journal of neurophysiology*, 92(4):2456–2467, 2004. 104, 105

[217] Austin M Ramsey, Ai-Hui Tang, Tara A LeGates, Xu-Zhuo Gou, Beatrice E Carbone, Scott M Thompson, Thomas Biederer, and Thomas A Blanpied. Subsynaptic positioning of ampars by lrrtm2 controls synaptic strength. *Science Advances*, 7(34):eabf3126, 2021. 104, 105

[218] Vladimir V Uteshev and Peter S Pennefather. A mathematical description of miniature postsynaptic current generation at central nervous system synapses. *Biophysical Journal*, 71(3):1256–1266, 1996. 104, 105

[219] Etsuko Tarusawa, Ko Matsui, Timotheus Budisantoso, Elek Molnár, Masahiko Watanabe, Minoru Matsui, Yugo Fukazawa, and Ryuichi Shigemoto. Input-specific intrasynaptic arrangements of ionotropic glutamate receptors and their impact on postsynaptic responses. *Journal of Neuroscience*, 29(41): 12896–12908, 2009. 105

[220] Shouto Yonekura, Alexandros Beskos, and Sumeetpal S Singh. Asymptotic analysis of model selection criteria for general hidden markov models. *Stochastic Processes and their Applications*, 132:164–191, 2021. 117, 148, 149

[221] Robert R Kerr, Anthony N Burkitt, Doreen A Thomas, Matthieu Gilson, and David B Grayden. Delay Selection by Spike-Timing-Dependent Plasticity in Recurrent Networks of Spiking Neurons Receiving Oscillatory Inputs. *PLOS Computational Biology*, 9(2):e1002897, February 2013. 119

[222] Luke Campagnola, Stephanie C Seeman, Thomas Chartrand, Lisa Kim, Alex Hoggarth, Clare Gamlin, Shinya Ito, Jessica Trinh, Pasha Davoudian, Cristina Radaelli, et al. Local connectivity and synaptic dynamics in mouse and human neocortex. *Science*, 375(6585):eabj5861, 2022. 121

[223] Taichi Murayama, Shoko Wakamiya, Eiji Aramaki, and Ryota Kobayashi. Modeling the spread of fake news on twitter. *Plos one*, 16(4):e0250419, 2021. 121

[224] Abed Ghanbari, Naixin Ren, Christian Keine, Carl Stoelzel, Bernhard Englitz, Harvey A Swadlow, and Ian H Stevenson. Modeling the short-term dynamics of in vivo excitatory spike transmission. *Journal of Neuroscience*, 40(21): 4185–4202, 2020. 121

Appendices

Appendix A

Fast-fitting methods and mean-field approximations

The following results have been presented during the Bernstein Conference 2020[1].

We first recall how a synapse with Short-Term Depression can be modeled as a Hidden-Markov Model (Figure A.1). Upon the arrival of a presynaptic action potential, vesicles from a pool of N independent release sites open with a probability p, each of these release events giving rise to a quantal current q. A synapse is thus described by its parameters N, p, and q [12], and the time constant τ_D governing short-term depression [17, 134]. The equations of the model are thus derived as follows:

$$p(\mathcal{D}|\theta) = \sum_{\mathbf{n},\mathbf{k}} p(\mathcal{D}, \mathbf{n}, \mathbf{k}|\theta)$$

with

$$p(\mathcal{D}, \mathbf{n}, \mathbf{k}|\theta) = p_\theta(e_1|k_1)p_\theta(k_1|n_1)p_\theta(n_1)\prod_{i=2}^{T} p_\theta(e_i|k_i)p_\theta(k_i|n_i)p_\theta(n_i|n_{i-1}, k_{i-1})$$

The state transition and emission probabilities of the model are defined as

$$[e_i|k_i] \sim \mathcal{N}(qk_i, \sigma^2)$$

$$[k_i|n_i] \sim \mathrm{Bin}(n_i, p)$$

and

$$[n_i - n_{i-1}|n_{i-1}, k_{i-1}] \sim Bin(N - (n_{i-1} - k_{i-1}), 1 - e^{-\Delta_i/\tau_D})$$

Our model is a Hidden Markov Process with observations $\mathcal{D}$ and hidden variables $\mathbf{n}$ and $\mathbf{k}$. The Baum-Welch algorithm is a forward-backward procedure classically used

[1] C. Gontier and J.-P. Pfister. Fast and online inference of synaptic parameters. 16$^{\text{th}}$ Bernstein Conference, Berlin, Germany. DOI: `https://doi.org/10.12751/nncn.bc2020.0029`

to compute the likelihood of a set of parameters $p(\mathcal{D}|\theta)$. However, its algorithmic complexity scales with N^4 and makes it unpractical for studying synapses with a large number of vesicles.

In Chapter 5, we propose the use the Nested Particle Filter [186] for online posterior computation. This algorithm is asymptotically exact and purely recursive, and thus allows to directly estimate the parameters as recordings are acquired. However, this weighted method requires a very large number of particles to provide low variance estimates, as the approximation error only decreases with the square root of the number of particles (see Figure D.3).

Here, we derive a mean-field approximation of the likelihood of a set of parameters. This likelihood computation is at the basis of most parameter inference methods, and our approximation (in which data are decorrelated) reduces its algorithmic complexity to N^2. Our mean-field approximation relies on two assumptions:

- We assume that data are decorrelated:

$$p(n_i) \approx p(n_i|n_{i-1}, k_{i-1})$$

  which allows to compute the likelihood of each data point e_i independently (see Figure A.2, left panel):

$$p(e_i|\theta) = \sum_{n_i,k_i} p(e_i|k_i)p(k_i|n_i)p(n_i)$$

- We assume that n_i are generated from a normal distribution: $n_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ where $\mu_i = \mathbb{E}(n_i)$ and $\sigma_i^2 = Var(n_i)$ are computed analytically using the laws of total expectation and variance (see Figure A.2, right panel).

We note n_i the number of filled vesicles just before spike i, $E(n_i)$ its expectation and $V(n_i)$ its variance.

$n_i = n_{i-1} - k_{i-1} + r$ with $r \sim \text{Bin}\,(N - n_{i-1} + k_{i-1}, I_i)$ and where I_i is defined as

$$I_i = 1 - \exp\left(-\frac{\Delta t_i}{\tau_D}\right)$$

According to the law of total variance:

$$V(n_i) = E(V(n_i|n_{i-1}, k_{i-1})) + V(E(n_i|n_{i-1}, k_{i-1}))$$

$$\implies V(n_i) = E(I_i(1-I_i)(N - n_{i-1} + k_{i-1})) + V(n_{i-1} - k_{i-1} + I_i(N - n_{i-1} + k_{i-1}))$$

$$\implies V(n_i) = I_i(1-I_i)(N - E(n_{i-1}) + pE(n_{i-1})) + V(I_i N + (1-I_i)(n_{i-1} - k_{i-1}))$$

$$\implies V(n_i) = I_i(1-I_i)(N - E(n_{i-1}) + pE(n_{i-1})) + (1-I_i)^2 V(n_{i-1} + k_{i-1})$$

The number of vesicles that do not release at a time i (assuming that n_i vesicles are filled) follows a binomial law as well: $n_i - k_i|n_i \sim \text{Bin}\,(n_i, 1-p)$

It thus follows from the law of total variance:

$$V(n_{i-1} - k_{i-1}) = E(V(n_{i-1} - k_{i-1}|n_{i-1})) + V(E(n_{i-1} - k_{i-1}|n_{i-1}))$$

$$\implies V(n_{i-1} - k_{i-1}) = E(p(1-p)n_{i-1}) + V((1-p)n_{i-1})$$
$$\implies V(n_{i-1} - k_{i-1}) = p(1-p)E(n_{i-1}) + (1-p)^2 V(n_{i-1})$$

Finally,

$$V(n_i) = I_i(1-I_i)(N - E(n_{i-1}) + pE(n_{i-1})) + (1-I_i)^2(p(1-p)E(n_{i-1}) + (1-p)^2 V(n_{i-1}))$$

Assuming that hidden states are well described by their first and second order moments, our mean-field approximation allows to significantly reduce the computation time of the likelihood function (Figure A.3) while remaining sufficiently exact for accurate parameters inference (Figure A.4).

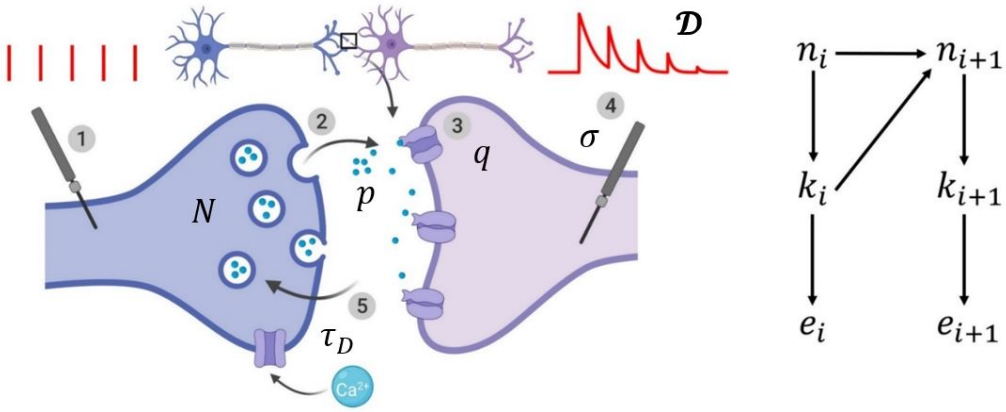


Figure A.1: **Modelization of a synapse with STD as a Hidden Markov Model**. The synapse is modelled as a Hidden Markov Model (HMM) with hidden states n_i (the number of vesicles in the filled state at the moment of spike i) and k_i (the number of vesicles among the n_i which will release the neurotransmitters following spike i). Between two successive spikes, the probability that a given vesicle is refilled within a time Δ_{t_i} is $I_i(\Delta_{t_i}) = 1 - \exp\left(-\frac{\Delta_{t_i}}{\tau_D}\right)$.

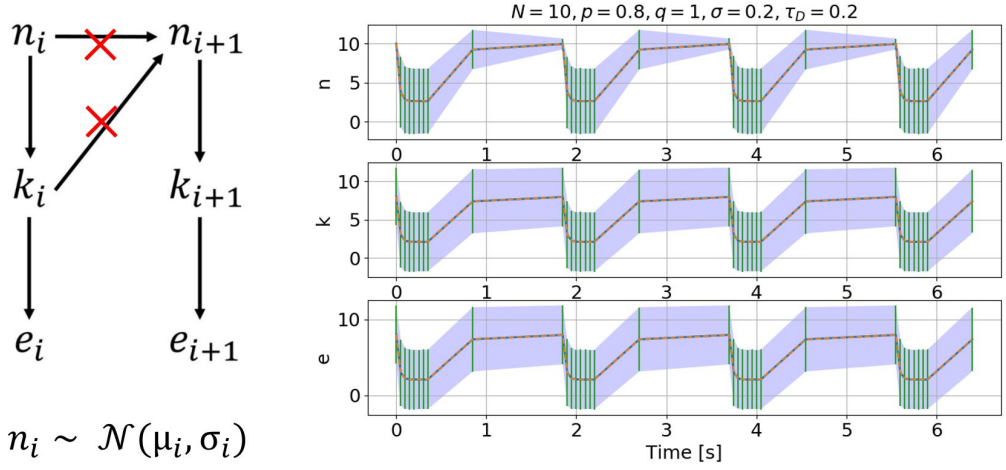


Figure A.2: **Mean-field approach to likelihood computation: theory**. Analytical mean and variance of hidden and observed variables (solid lines) computed using the mean-field approximation are compared to empirical values (dashed line and shaded area) from 1000 synthetic data sets.

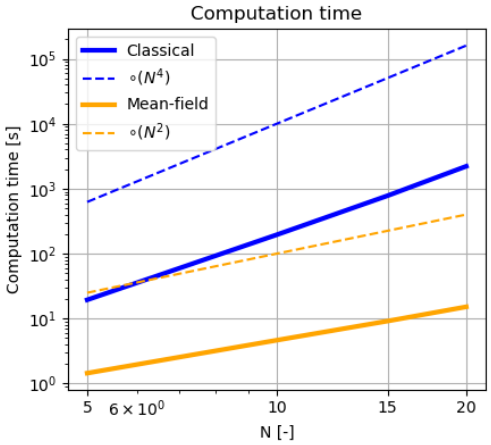


Figure A.3: **Mean-field approach to likelihood computation: results (computation time)**. Mean computation time (averaged over 1000 synthetic data sets) of the likelihood of a set of 200 data points.

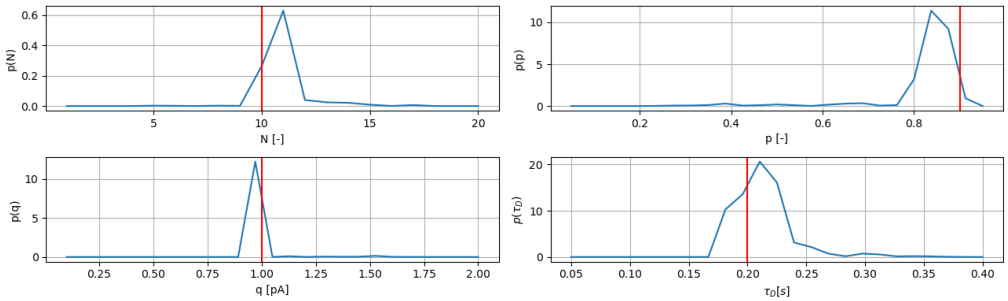


Figure A.4: **Mean-field approach to likelihood computation: results (posterior)**. Posterior distributions of the parameters computed using the Metropolis-Hastings algorithm. Likelihood computation was performed using the mean-field approximation. Ground-truth values used to generate the EPSCs are displayed as red vertical lines.

Appendix B

Model selection for correlated data

In this brief chapter we address the question of how to perform model selection (and especially how to use the BIC) for models with correlated data, i.e. models for which observations are not i.i.d. As explained in Section 4.3.6, the likelihood of observing data $\mathcal{D}$ given a model $\mathcal{M}$ (i.e. the model evidence) is given by

$$p(\mathcal{D}|\mathcal{M}) = \int d\theta p(\mathcal{D}|\theta, \mathcal{M})\pi(\theta|\mathcal{M})d\theta \tag{B.1}$$

where $\pi(\theta|\mathcal{M})$ is the prior for the parameters of $\mathcal{M}$. As this integral is often intractable, an interesting approximation is called the Bayesian Information Criterion (BIC) [55]:

$$BIC_{\mathcal{M}}(\mathcal{D}) = -2\log p(\mathcal{D}|\hat{\theta}, \mathcal{M}) + k_{\mathcal{M}}\log(T) \approx -2\log p(\mathcal{D}|\mathcal{M}) \tag{B.2}$$

which derivation is detailed in Section 4.6.1. However, the approximation $BIC_{\mathcal{M}}(\mathcal{D}) \approx -2\log p(\mathcal{D}|\mathcal{M})$ is only valid under the hypothesis that observations are i.i.d., which does not hold for for HMM[1]: in the general case, $p(y_{t-1}, y_t) \neq p(y_{t-1})p(y_t)$ (Figure B.1).

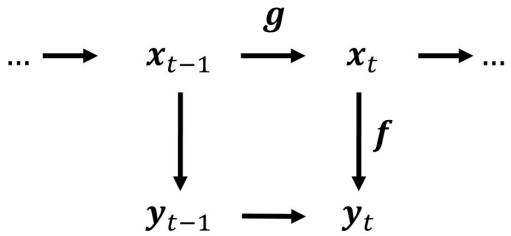


Figure B.1: **Observations in HMM are not i.i.d.** Illustration of a HMM with hidden state x_t, observation y_t, state transition g and emission probability f. In the general case, observations are not independent, i.e. $p(y_{t-1}, y_t) \neq p(y_{t-1})p(y_t)$.

If data are correlated, we are left with the following approximation, which does not simplify in the general case:

[1]The same i.i.d. requirement also holds for other model selection criteria, such as the AIC.

$$-2\log p(\mathcal{D}|\mathcal{M}) \approx -2\log \pi(\hat{\theta}|\mathcal{M}) - 2\log p(\mathcal{D}|\hat{\theta},\mathcal{M}) - k_{\mathcal{M}}\log(2\pi) + \log(|-H(\hat{\theta})|) \quad \text{(B.3)}$$

Different solutions can be proposed to approximate (B.3). Firstly, data can be "decorrelated", for instance by approximating the likelihood through a mean-field approach (i.e. $p(y_{t-1}, y_t) \approx p(y_{t-1})p(y_t)$), which allows to use the classical definition of the BIC for i.i.d. data. However, depending on the nature of the studied system, neglecting correlations in data can be a crude approximation and lead to biased estimates of the likelihood.

Secondly, a numerical approximation of the Hessian $H(\hat{\theta})$ can be obtained through MCMC methods, event without an explicit expression for the gradient, see for example [158].

Finally, if the likelihood of the model can be computed exactly (using for instance the Baum-Welch algorithm), the Hessian matrix can be computed analytically (which is the approach that we used in Section 4.6.3). But computing this matrix can be cumbersome (for a model with $k_{\mathcal{M}}$ free parameters, there are $k_{\mathcal{M}}^2$ double derivatives to compute). For more complicated models, such a computation can be intractable.

Posterior to our publication on practical identifiability (Chapter 4), a very insightful paper, written by Shouto Yonekura, Alexandros Beskosa, and Sumeetpal S. Singh, was published in *Stochastic Processes and their Applications*[2]. In this paper [220], the authors also address the validity of classically used Model Selection Criteria (including the BIC) for general Hidden Markov Models. Their main result is that, under regularity conditions (namely, that the observation generating process is strongly stationary and ergodic), the BIC remains strongly consistent for HMMs. More specifically, they show that the determinant of the Hessian matrix grows linearly with the number of observations T:

$$\lim_{T\to+\infty} |H(\hat{\theta}) - T\mathcal{I}_{\hat{\theta}}| = 0 \quad \text{(B.4)}$$

where $\mathcal{I}_{\hat{\theta}}$ is a non-singular matrix (see Section 3 in [220] for a detailed discussion). The same approximation as in Section 4.6.1 in the limit of large T can then be made in order to obtain $BIC_{\mathcal{M}}(\mathcal{D}) \approx -2\log p(\mathcal{D}|\mathcal{M})$. This result has already been used for model selection with HMMs[3].

This result only applies to strongly stationary and ergodic observation generating processes, i.e. to HMMs for which the state transition and emission probabilities (functions f and g in Figure B.1) are constant. It thus does not generalize to the more general class of Input-Output Hidden Markov Models (IO-HMMs, also called GLM-HMMs in neuroscience [190]), in which the state transition probability at time t depends on an external input ψ_t. For instance, the state transition in our model of synapse is not stationary, but depends on the ISI ψ_t (see Section 5.6). If the input is constant (i.e. when ψ_t does not depend on t), the model corresponds to a classical stationary HMM. However, in most physiological experiments, ISIs are not

[2] Yonekura, S., Beskos, A., and Singh, S. S. (2021). Asymptotic analysis of model selection criteria for general hidden Markov models. Stochastic Processes and their Applications, 132, 164-191. doi: https://doi.org/10.1016/j.spa.2020.10.006

[3] Narula et al. In preparation.

constant but either drawn from a distribution (e.g. an exponential distribution [159]) or defined in a cyclic pattern [20] consisting of a high-frequency tetanic stimulation phase followed by recovery spikes.

We argue that, if the model generating ψ_t is itself stationary, the results proposed in [220] could still hold. We verify empirically this intuition using numerical simulations (Figure B.2). Surrogate data are generated from a binomial model with STD and with ground-truth parameters $N = 5$, $p = 0.7$, $q = 1$, $\sigma = 0.2$, and $\tau_D = 0.2$. Three different processes for generating the ISIs are proposed:

- **Constant protocol**: all ISIs are constant and equal to 0.05s;
- **Exponential protocol**: ISIs are drawn from a (stationary) exponential distribution with mean $\lambda = 0.05$s;
- **Cyclic protocol**: a train of ISIs, consisting of 6 ISIs of 0.05s and 4 recovery ISIs of 0.075s, 0.1s, 0.5s, and 5s, is repeated.

The second derivative of the likelihood function $p(e_{1:t}|\tau_D)$ is plotted for these 3 stationary protocols as the number of generated observations t increases: in all cases, we can see a linear increase with the number of observations. We thus argue that the result in [220] can be generalized to the broader class of IO-HMM, given that the process generating the inputs ψ_t is itself stationary.

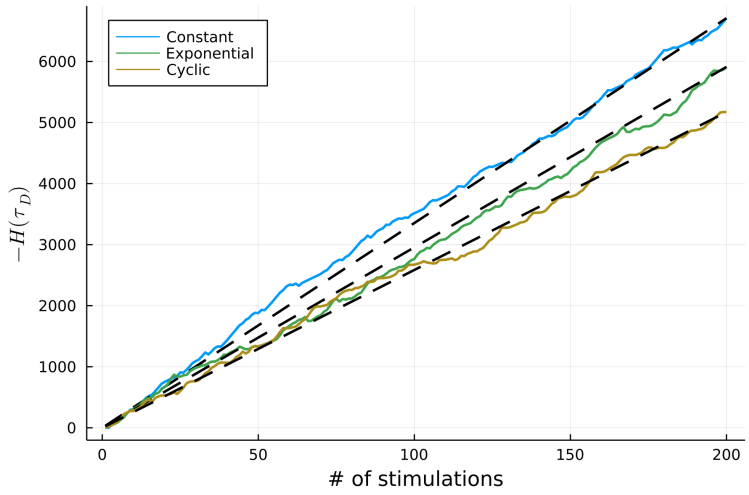


Figure B.2: **Applicability of the BIC to IO-HMM**. Surrogate data are generated from a binomial model with STD and with ground-truth parameters $N = 5$, $p = 0.7$, $q = 1$, $\sigma = 0.2$, and $\tau_D = 0.2$ using three different stimulation protocols (Constant, Exponential, Cyclic). Lines show the second derivative of the likelihood function $p(e_{1:t}|\tau_D)$, denoted $-H(\tau_D)$. A nearly-linear increase of this quantity with t can be observed.

Appendix C

Occam's razor for nested models

An interesting result in line with our proposed definition of practical model identifiability (Chapter 4) would be to prove the following inequality:

$$\forall \theta_0, \langle \log p(\mathcal{D}|\mathcal{M}_0) \rangle_{p(\mathcal{D}|\theta_0,\mathcal{M}_0)} \geq \langle \log p(\mathcal{D}|\mathcal{M}_1) \rangle_{p(\mathcal{D}|\theta_0,\mathcal{M}_0)}$$

where $\mathcal{M}_0$ and $\mathcal{M}_1$ are competing models with the former being nested in the latter. This intuitively corresponds to an application of Occam's razor (a simpler and true model will be favored over a more complicated one), but lacks a formal proof.

Proposition: Let $\mathcal{M}_0$ and $\mathcal{M}_1$ be two nested models such that $\mathcal{M}_0 \preceq \mathcal{M}_1$. We denote Θ_0 and Θ_1 the space of possible parameters for $\mathcal{M}_0$ and $\mathcal{M}_1$, with $\Theta_0 \subset \Theta_1$. If data generated from $\mathcal{M}_0$ and $\mathcal{M}_1$ are i.i.d., then the following inequality holds $\forall \theta_0^* \in \Theta_0$:

$$\langle \log p(\mathcal{D}|\mathcal{M}_0) \rangle_{p(\mathcal{D}|\theta_0^*,\mathcal{M}_0)} \geq \langle \log p(\mathcal{D}|\mathcal{M}_1) \rangle_{p(\mathcal{D}|\theta_0^*,\mathcal{M}_0)} \tag{C.1}$$

If data are not i.i.d., a sufficient condition for the inequality to hold is

$$k_{\mathcal{M}_0} \log(2\pi) - \sum_{i=1}^{k_{\mathcal{M}_0}} \langle \log(\lambda_i^0) \rangle_{p(\mathcal{D}|\theta_0^*,\mathcal{M}_0)} \geq k_{\mathcal{M}_1} \log(2\pi) - \sum_{i=1}^{k_{\mathcal{M}_1}} \langle \log(\lambda_i^1) \rangle_{p(\mathcal{D}|\theta_0^*,\mathcal{M}_0)} \tag{C.2}$$

where

$k_{\mathcal{M}_0}$ and $k_{\mathcal{M}_1}$ are the number of independent parameters of $\mathcal{M}_0$ and $\mathcal{M}_1$;

$H_0(\hat{\theta}_0)$ and $H_1(\hat{\theta}_1)$ are the Hessian matrices of the log-likelihoods $p(\mathcal{D}|\theta_0, \mathcal{M}_0)$ and $p(\mathcal{D}|\theta_1, \mathcal{M}_1)$ expressed at their respective MLEs;

$\{\lambda_i^0\}_{1 \leq i \leq k_{\mathcal{M}_0}}$ and $\{\lambda_i^1\}_{1 \leq i \leq k_{\mathcal{M}_1}}$ are the respective eigenvalues of $-H_0(\hat{\theta}_0)$ and $-H_1(\hat{\theta}_1)$.

Reasoning: using the same approximation as in the derivation of the BIC (Section 4.6.1) for $p(\mathcal{D}|\mathcal{M}_0)$ and $p(\mathcal{D}|\mathcal{M}_1)$ yields

$$\log p(\mathcal{D}|\mathcal{M}_0) = \log p(\mathcal{D}|\hat{\theta}_0, \mathcal{M}_0) + \log \pi(\hat{\theta}_0|\mathcal{M}_0) + \frac{k_{\mathcal{M}_0}}{2}\log(2\pi) - \frac{1}{2}\log(|-H_0(\hat{\theta}_0)|) \tag{C.3}$$

$$\log p(\mathcal{D}|\mathcal{M}_1) = \log p(\mathcal{D}|\hat{\theta}_1, \mathcal{M}_1) + \log \pi(\hat{\theta}_1|\mathcal{M}_1) + \frac{k_{\mathcal{M}_1}}{2}\log(2\pi) - \frac{1}{2}\log(|-H_1(\hat{\theta}_1)|) \tag{C.4}$$

Both quantities then need to be averaged over $\langle\cdot\rangle_{p(\mathcal{D}|\theta_0^*,\mathcal{M}_0)}$. Assuming

$$\langle \log p(\mathcal{D}|\hat{\theta}_0, \mathcal{M}_0)\rangle_{p(\mathcal{D}|\theta_0^*,\mathcal{M}_0)} \approx \langle \log p(\mathcal{D}|\theta_0^*, \mathcal{M}_0)\rangle_{p(\mathcal{D}|\theta_0^*,\mathcal{M}_0)} \tag{C.5}$$

(i.e. that the maximum likelihood estimator $\hat{\theta}_0$ will be close to the true value θ_0^* from which data were generated) yields

$$\langle \log p(\mathcal{D}|\hat{\theta}_0, \mathcal{M}_0)\rangle_{p(\mathcal{D}|\theta_0^*,\mathcal{M}_0)} \geq \langle \log p(\mathcal{D}|\hat{\theta}_1, \mathcal{M}_1)\rangle_{p(\mathcal{D}|\theta_0^*,\mathcal{M}_0)}$$

(under Gibbs's inequality). Furthermore, $k_{\mathcal{M}_0} \leq k_{\mathcal{M}_1}$ yields $\pi(\hat{\theta}_0|\mathcal{M}_0) \geq \pi(\hat{\theta}_0|\mathcal{M}_1)$ (these quantities do not depend on $\mathcal{D}$). The inequality is thus met for the first two terms on the right-hand side.

For the last two terms, if data are i.i.d. and if the number of data points T in $\mathcal{D}$ is sufficiently large, the same approximation as in the derivation of the BIC can be made:

$$\frac{k_{\mathcal{M}}}{2}\log(2\pi) - \frac{1}{2}\log(|-H(\hat{\theta})|) \approx -\frac{k_{\mathcal{M}}}{2}\log(T)$$

Since $k_{\mathcal{M}_0} \leq k_{\mathcal{M}_1}$, the inequality thus holds if data generated from $\mathcal{M}_0$ and $\mathcal{M}_1$ are i.i.d.

If data are correlated, the above approximation does not hold. However, the determinant of the Hessian (which is a symmetric matrix) can be written as the product of the eigenvalues, which finally leads to the necessary condition. This inequality can also be seen as a more general version of a result presented in the following paper using less stringent approximations [65].

Appendix D

Video examples of different filtering techniques

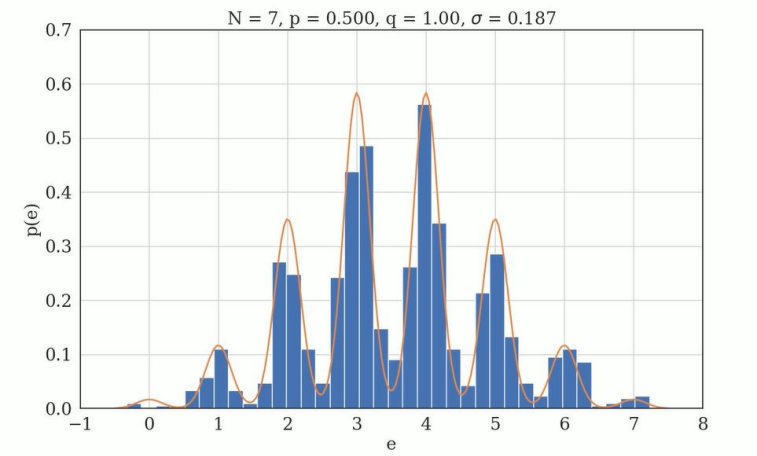


Figure D.1: **Illustration of Maximum Likelihood Estimation (MLE)**. In MLE, the parameters θ of a model $\mathcal{M}$ are fitted to a set of observations $\mathcal{D}$ to maximize the likelihood $p(\mathcal{D}|\theta, \mathcal{M})$. Computing $\arg\max_\theta p(\mathcal{D}|\theta, \mathcal{M})$ can be done using different optimization schemes, e.g. the Expectation-Maximization algorithm [20]. Here, data were generated from the binomial model with ground-truth parameters θ^* ($N = 7$, $p = 0.5$, $q = 1$, $\sigma = 0.2$). Their histogram is shown in blue. The theoretical distribution $p(\mathcal{D}|\theta, \mathcal{M})$ is shown in orange for different values of θ displayed on top. The animation shows how the theoretical distributions provides a better fit to the data as θ gets closer to θ^*. Link to the video: `https://youtu.be/43QKYkAVKl4`

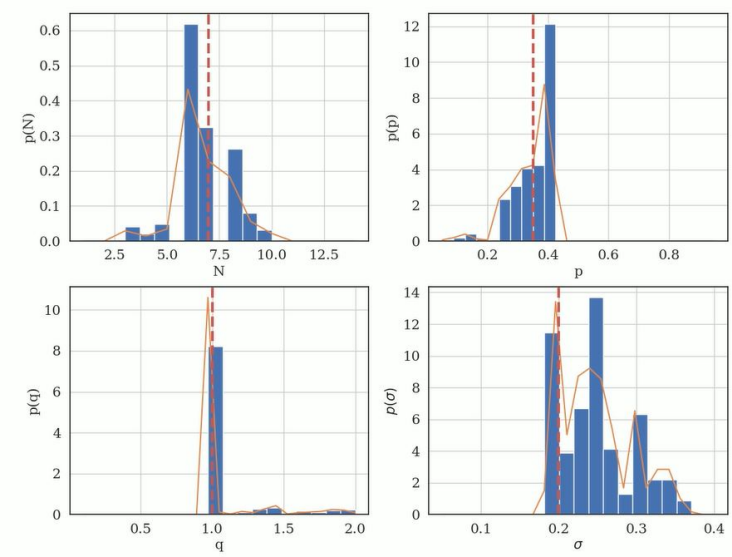


Figure D.2: **Illustration of the Metropolis-Hastings (MH) algorithm**. In Bayesian inference, the goal is to infer the full posterior distribution of the parameters $p(\theta|\mathcal{D}, \mathcal{M}) \propto p(\mathcal{D}|\theta, \mathcal{M})$ (as opposed to MLE, where only a point-based estimate $\theta^* = \arg\max_\theta p(\mathcal{D}|\theta, \mathcal{M})$ is computed). The MH algorithm is a rejection sampling scheme which allows to compute such posteriors [5]. Here, data were generated from the binomial model with ground-truth parameters $N = 7$, $p = 0.35$, $q = 1$, and $\sigma = 0.2$ (vertical red dashed lines). As samples (blue histogram) accumulate, they shape the posterior $p(\theta|\mathcal{D}, \mathcal{M})$. Link to the video: `https://youtu.be/pBOS2pkw-jk`

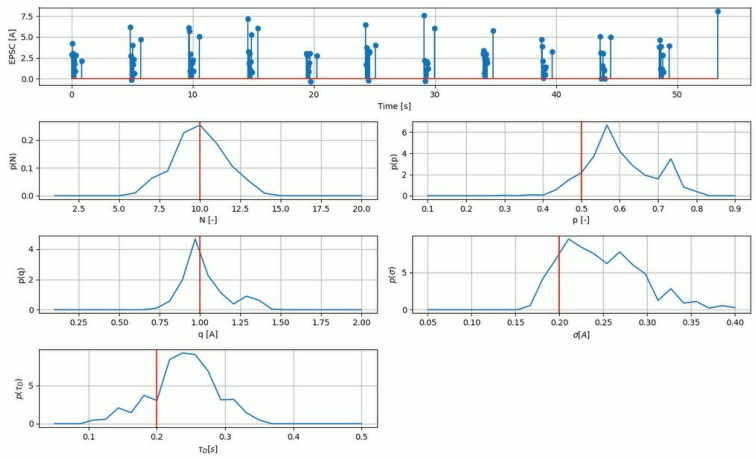


Figure D.3: **Illustration of the Nested Particle Filter (NPF)**. The NPF [186] is a non-linear particle filtering algorithm used to infer parameters of HMMs. It is asymptotically exact and purely recursive, and thus allows to directly estimate the distribution of parameters as recordings are acquired. The NPF relies on two nested layers of particles to approximate the posterior distributions of both the static parameters of the model and of its hidden states. A first outer filter with M_{out} particles is used to compute the posterior distribution of parameters $p(\theta|\mathcal{D})$, and for each of these particles, an inner filter with M_{in} particles is used to estimate the corresponding hidden states (so that the total number of particles in the system is $M_{\text{out}} \times M_{\text{in}}$). After each new observation, these particles are resampled based on their respective likelihoods. Its implementation for synaptic characterization is detailed in Section 5.4. Link to the video: `https://youtu.be/OPGEyayhxJI`. Upper panel: train of synthetic EPSCs generated from a model of synapse with short-term depression. Lower panels: posterior distributions of the parameters computed online using the nested particle filter. Ground-truth values used to generate the EPSCs are displayed as red vertical lines.

Declaration of Originality

Last name, first name:

Matriculation number:

I hereby declare that this thesis represents my original work and that I have used no other sources except as noted by citations.
All data, tables, figures and text citations which have been reproduced from any other source, including the internet, have been explicitly acknowledged as such.
I am aware that in case of non-compliance, the Senate is entitled to withdraw the doctorate degree awarded to me on the basis of the present thesis, in accordance with the "Statut der Universität Bern (Universitätsstatut; UniSt)", Art. 69, of 7 June 2011.

Place, date

Signature