UNIVERSITY OF CALIFORNIA SAN DIEGO

Models of biological learning across different spatial scales

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Physics

by

Bin Wang

Committee in charge:

Professor Johnatan Aljadeff, Chair
Professor David Kleinfeld, Co-Chair
Professor Marcus Benna
Professor Nigel Goldenfeld
Professor Duygu Kuzum

2023

The Dissertation of Bin Wang is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost, I am truly grateful to my advisor Yonatan Aljadeff, who led me into the wonderful realm of learning in neuroscience when uncertainty clouded my path towards a PhD. Time and again, Yonatan has consistently offered me his invaluable advice and unwavering support. I cannot remember how many times when I was stuck in the long dark night of my project, Yonatan will always be there and magically pinpoint the right direction to find the light and guide me to move forward. I am continually in awe of by Yonatan's incredible insights to identify the key ingredients and intuitions behind complex and sometimes messy neuroscience phenomena and to be full of fantastic ideas to elegantly connect simple models to experimental data. Such an research style has profoundly influence my work, constantly reminding me to focus on the essence of the problem and keep the experimental data in mind. Beyond directly helping my research, Yonatan has been an tremendous mentor in nurturing my evolution as a scientist. With unwavering generosity, he has dedicated countless hours to provide invaluable advice during my presentation and talk preparations, to connect me to collaborate with leading experimentalists worldwide, and to extend guidance on many other facets of my career development. My endeavors of this PhD would have been an elusive dream without his pivotal role.

I am also very grateful to receive advice and guidance from other mentors and faculty members throughout my PhD journey. I had the wonderful learning experiences to teach with David Kleinfeld in the course on biophysical basis of neurons and networks. Through numerous inspiring lectures and engaging conversations with David, I gained a much deeper understanding and appreciation for the advancements, potential, and limitations of contemporary neuroscience experimental techniques. His scientific passion and personal warmth has been a constant source of motivation for me to keep exploring and pursuing the mysteries of the brain. I would also like to thank Olga Dudko for introducing me the mathematical and quantitative approach to neuroscience phenomena and for her consistent support to my project. Olga's insistence on discerning general principles amidst diverse and distinct biological systems has profoundly

shaped my research perspective. I have also benefited a lot from many stimulating discussions with Marcus Benna regarding my thesis projects and theoretical neuroscience in general. I would also like to thank Nigel Goldenfeld and Duygu Kuzum for their thoughtful feedback to the works included here. I am also fortunate to learn about the bewildering world of probability theory from Todd Kemp during his numerous enlightening lectures and discussions and to gain a deeper understanding of the mathematical tools from statistical physics from Haim Sompolinsky during the summer school at Woods Hole. I also want to express my special thank to our experimental collaborators, Nick Audette and David Schneider at NYU. Their insights and patience in clarifying intricate experimental details and the connections to our models have been invaluable.

During my PhD journey, I've been fortunate to forge meaningful friendships with a remarkable group of supportive peers, who have truly enriched this academic adventure. In particular, I want to thank the members from Aljadeff's group, Shiva Azizpour Lindi, Yu Terada and Palka Puri, with whom we have had many lovely conversations and who have created an incredibly intellectual environment. I would also like to thank Haixiao Wang especially, who has been my roommate for five years. We frequently hang out together for Chinese food and had many stimulating conversations about mathematics and academics. I also enjoy lots of illuminating discussions with my peers in the computational neuroscience and biological physics community at UCSD. In particular, I want to thank Xiang Ji, Johnney Li, Ji-An Li, Pengcen Jiang, Tynan Kennedy, Haoliang Wang, Huanqiu Zhang, Rui Liu, Pantong Yao, Yusi Chen, Dhruv Kohli, Meg Robinson, Yuru Song, Stella Yuan, for their friendship and support.

Finally, I am deeply grateful to my beloved parents, Haifang Yang and Suxu Wang for their unconditional love, encouragement, and caring. I'm also grateful to my partner, Zixi Lin, for her unwavering support and for brightening my PhD journey. This thesis would not have been possible without your support.

Chapter 2, in full, is a reprint of the material as it appears in Wang and Dudko. "A theory of synaptic transmission." Elife 10 (2021): e73585. The dissertation author was the primary

author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in Wang and Aljadeff. ”Multiplicative Shot-Noise: A New Route to Stability of Plastic Networks.” Physical Review Letters 129.6 (2022): 068101. The dissertation author was the primary author of this paper.

Chapter 4, in part is currently being prepared for submission for publication of the material. Wang, Audette, Schneider and Aljadeff. ”Desegregation of neural predictive processing”. The dissertation author was the primary author of this material.

VITA

2018 Bachelor of Sciences in Integrated Science, Peking University, China

2023 Doctor of Philosophy in Physics, University of California San Diego

PUBLICATIONS

Bin Wang, Nicholas J Audette, David M Schneider and Johnatan Aljadeff. ”Desegregation of neural predictive processing” (In preparation).

Bin Wang and Johnatan Aljadeff. ”Multiplicative Shot-Noise: A New Route to Stability of Plastic Networks.” Physical Review Letters 129.6 (2022): 068101.

Bin Wang and Olga K. Dudko. ”A theory of synaptic transmission.” Elife 10 (2021): e73585.

ABSTRACT OF THE DISSERTATION

Models of biological learning across different spatial scales

by

Bin Wang

Doctor of Philosophy in Physics

University of California San Diego, 2023

Professor Johnatan Aljadeff, Chair
Professor David Kleinfeld, Co-Chair

A hallmark of the brain is its capacity for learning, which broadly refers to the animal's ability to constantly adapt to the changing environments by adjusting its behaviors. Learning in the brain happens across a wide range of spatiotemporal scales. Here I will present three modeling works that link biological mechanisms of learning across different spatial scales, including linking molecular mechanisms to cellular function of synaptic transmission, the effects of synaptic plasticity on network dynamics, and circuit mechanisms for multi-sensory perception. Beyond their specific contexts, my dissertation research will provide general methods and tools to bridge learning mechanisms across scales and elucidate principles governing the brain's

capability of learning, one of the unique features of intelligence.

Chapter 1

Introduction: Seeking for Principles of Learning in the Brain

1.1 The phenomena of learning

Humans, like many other animals, possess an intrinsic ability to adjust and adapt their behaviors in response to changing environment, based upon past experiences. Such experience-dependent changes in behavior are commonly termed *learning*. Besides behavioral changes, learning is often accompanied with the acquisition of new knowledge, skills, values and the formation or removal of memories. Some learning experiences are rapid and can be induced by a single event (e.g. learning perils of heats by touching a hot stove), while others can be slow and evolve gradually through repeated experiences (e.g. learning a new language). Learning is a fundamental function of the nervous system, integral to our perception, reasoning, thinking, and interaction with the surrounding world.

Learning is pervasive in our lives. For instance, when traveling to a new city, one may initially rely heavily on maps or seek guidance from locals to locate essential places related to daily life, like workplaces, residences, or entertainment venues. As one traverses the city's streets and pathways repeatedly, the brain is able to learn a spatial map of the new environment, which then allows us to navigate to the relevant locations independent of external guidance, or to find new places based on those places that we are familiar with. Similarly, when learning a new sport, we may begin by mimicking specific movements. But over time, through consistent practice,

these movements become second nature, almost reflexive. It is as if the brain has encoded an internal program, seamlessly initiating these learned action sequences whenever prompted. As a final example, the journey to a PhD stands for a more complex form of learning. It begins with absorbing established knowledge but culminates in contributing novel insights and expanding the collective knowledge base. It is remarkable that the human brain can undergo such profound transformations in a span of a few years and successfully learn to perform such a sophisticated task.

Although this dissertation primarily focuses on learning phenomena within the brain, brain-based learning can be viewed as a special form of adaptive behaviors in a broader context. These adaptive behaviors can be observed even in organisms without a nervous system and are crucial for species survival in the ever-evolving environmental changes. Moreover, the observation that animals can learn from experience has inspired recent advances in artificial intelligence, where machines are built to learn from scratch to perform specific tasks. Yet in many real-life scenarios, there is still a large gap between the performance of state-of-the-art artificial systems and the brain. Thus, diving deeper into the brain's learning mechanisms offers a twofold benefit: it advances our understanding of the operating principles of nervous systems and broader adaptive principles in biology, and also paves the way for designing artificial systems that rival, or even surpass, human performance levels.

This dissertation concerns modeling biological mechanisms underlying learning in the brain. The overarching goal is to uncover principles that govern the learning process of nervous systems, which can be formulated mathematically and can be applied to explain a plethora of learning phenomena. To set the stage for the results presented in this dissertation, the subsequent two sections will dive into some backgrounds on mechanistic and normative approaches to model learning phenomena in the brain. The closing section of this chapter will contain a summary of the main results of the dissertation.

1.2 A mechanistic viewpoint of learning

This section reviews biological mechanisms of learning on the molecular, cellular and circuit levels. Particular emphasis is placed on those mechanisms that provide context for the results presented in subsequent chapters.

Most of the learning processes in biological systems rely on the nervous system. Within the nervous system, neurons (or nerve cells) are the fundamental units: they sense changes in environment, communicate these changes to other neurons, and control the body's response to these sensations [21]. The communication between neurons is usually via synapses (or synaptic junctions), which connect the axon of the presynaptic neuron to the soma, dendrite or axon of the postsynaptic neuron. A large body of experimental evidence [107, 106, 153] has suggested that behavioral changes induced by learning are closely linked to the modification of various properties of synapses. The mechanisms that induce synaptic changes are collectively called synaptic plasticity [175]. Synaptic changes caused by synaptic plasticity result in changes of neural responses and influence the behaviors of the animals. In order to understand learning, it is necessary to identify what properties of a synapse can be tuned by plasticity.

Most of the synapses in mature animals are chemical synapses, where an electrical signal (action potential) of the presynaptic neuron triggers the release of neurotransmitters into the synaptic cleft and results in the response of the postsynaptic neuron. At most chemical synapses, neurotransmitters are packed into synaptic vesicles at specific zones at the presynaptic axon terminals, known as active zones. When the presynaptic neuron emits an action potential, it triggers calcium influx into the active zones. The calcium ions bind protein complexes residing on the membrane of synaptic vesicles, called SNARE proteins [104, 35]. Following binding of calcium, the SNARE complex undergoes configuration change and pulls the vesicle membrane towards the presynaptic cell membrane, inducing membrane fusion (or exocytosis). The properties of SNARE proteins and their effects on neurotransmitter release have been studied in numerous modeling and experimental works before [104, 224, 35, 171]. These works have

shown that the transmission properties of SNARE complex can vary dramatically across different synaptic types, brain regions and animal species. Even inside the same synapse, structures and components of SNARE complex can be modulated by various regulatory molecules to contribute to synaptic plasticity on the presynaptic side.

Besides SNARE proteins, other properties of the synapses can also change in response to the history of synaptic activity, such as the number of synaptic vesicles at the active zones, the magnitude of calcium concentration and so on. The variations of these properties all contribute to synaptic plasticity, especially to short-term plasticity [196, 99]. Given the immense diversity of molecular components at synapses, it is especially challenging to formulate a quantitative framework that can capture such heterogeneity and be used to shed light on its effects on various synaptic functions, including synaptic plasticity.

The synaptic changes induced by short-term plasticity usually last less than a few minutes [196]. Yet behavioral changes induced by learning can last for much longer times, which usually requires long-term change of synaptic properties through long-term plasticity. One primary mechanism of long-term synaptic plasticity also relies on the intracellular calcium. The depolarization/hyperpolarization of postsynaptic membrane following neurotransmitter release induces calcium influx into postsynaptic cytoplasm. Such a calcium influx triggers a cascade of molecular signaling pathways at the postsynaptic side, causes synthesis or degradation of synaptic proteins and eventually induces long-term changes of synaptic properties. A comprehensive description of the molecular pathways underlying long-term plasticity has been advanced in recent years [128, 150]. But these models are typically too complex to be used for understanding learning in neural circuits.

A different line of research has directly investigated the effect of plasticity induced at spike-level, called spike-timing dependent plasticity (STDP) [25, 60]. These studies measure the change of excitatory postsynaptic potential (EPSP) when stimulating the neurons by spike-pairs [24], triplets [75] and other patterns of spiking activities [60]. Recent works [79, 97] combined these phenomenological models for STDP to more detailed, biophysical models of synaptic

plasticity [214, 78] and proposed the calcium-based STDP model. Yet it remains unknown how the nonlinearity in the calcium-based model, measured at physiological conditions, affects the dynamics and structures at the network level. These results will shed light on the diverse place field dynamics as observed in experiments [58].

Synaptic plasticity on the network level also plays important roles in the brain's sensory processing. One ultimate goal of neuroscience is to understand how the brain internally represents the external world based on the sensory input signals. On the cognitive level, an effective mental representation of the world helps the animal to take correct actions under different environmental conditions. Early works have suggested that neurons in the brain encode specific features of the external sensory signals, called the "receptive field" of the neuron. Following this idea, neurons with different receptive fields have been identified in various brain regions, such as orientation neurons in visual cortex [94], place cells in hippocampus [181], direction neurons in motor regions [72] and so on. The receptive field idea typically highlights that each neuron detects features of the sensory stimuli that are currently presented. However, more and more recent works have shown that many neurons in the sensory regions encode features of those stimuli "expected" from past experiences rather than the ones that are actually presented [259, 210]. These observations not only indicate that sensory representations in the brain is experience-dependent but also show that the brain can compute an expectation of the world internally rather than just passively sense it. These ideas form the basis of predictive processing hypothesis [114].

The predictive processing hypothesis asserts that the brain is constantly generating and updating an internal model of the external world. As the internal model is updated based on past experiences, predictive processing is a prototypical example of learning in sensory processing. Early experimental evidence for predictive processing is mainly based on sensory representation of a single stimulus [93], like a image or sound. Recently, many works have identified predictive signals in the brain for associative learning of multisensory stimuli [259, 70]. Current research is starting to elucidate how learning changes the predictive representations formed in the neural circuits and how these are influenced by noise and structures in the sensory signals.

1.3 A normative viewpoint of learning

In theoretical neuroscience, the normative viewpoint emphasizes the brain's functions and computations [45]. This approach offers a conceptual framework to decipher the intricate data from large-scale neural recordings and offers a potential route to reveal "design principles" of nervous systems. Within the realm of learning, this viewpoint seeks to address questions like: what is learning good for? what specific brain function does the learning support?

From a functional perspective, it is useful to distinguish the two separate systems: the brain and the external world. The brain processes sensory inputs, form its internal representations and acts upon the external world. In this picture, learning changes the brain's representations and refines the actions it takes. Through learning, the brain can form a better representation of the surrounding environments and guide the animal to take better actions to survive in severe environments. The normative viewpoint posits that learning changes the state of the brain in such a way to optimize its functions (e.g. performance on specific task). Based on this, learning can be reduced to an optimization problem, i.e. identifying the optimal solution (within a model class) for task performance.

Typically, two major classes of problems arise: statistical problem and computational problem. The statistical problem is related to how to achieve successful learning with noisy and partially observed sensory inputs. The computational problem focus on how to effectively find the optimal solution during learning. Yet, there are also functional questions that do not neatly fit in the two classes.

As an example, sparse coding has been suggested as a normative principle for how sensory system encodes external stimuli [64, 200]. Suppose the external sensory signal is denoted as $x(t) \in \mathbb{R}$. We use neural activity $\boldsymbol{r}(t) \in \mathbb{R}^N$ to describe the states of a sensory region in the brain. Sparse coding asserts that the neural activity minimizes the following objective function

$$E(\boldsymbol{r}(t)) = \frac{1}{2}\left[x(t) - \boldsymbol{w} \cdot \boldsymbol{r}(t)\right]^2 + F(\boldsymbol{r}(t)). \tag{1.1}$$

where the first term is an error with respect to a linear readout of the neural activity and the second term $F(\boldsymbol{r})$ forces sparsity. Here the task is to reconstruct the sensory signal $x(t)$, whose performance is measured by the first term. From statistical perspective, the second term serves as an regularization to avoid overfitting.

For a static sensory signal $x(t) = x$, the optimal neural response is given by

$$\boldsymbol{r}^* = \underset{\boldsymbol{r}}{\operatorname{argmin}} \left[\frac{1}{2} (x - \boldsymbol{w} \cdot \boldsymbol{r})^2 + F(\boldsymbol{r}) \right]. \tag{1.2}$$

From computational perspective, such an optimal solution of the brain state can be found by a standard recurrent neural network with nonlinear activation function [200]. It turns out that these mathematical formalism can be extended to understand the predictive processing of multimodal, high-dimensional sensory signals.

1.4 Outline: bridging learning mechanisms across spatial scales

As many biological systems, the brain is notably a multiscale system and spans across different spatial scales, from molecular and cellular level up to circuit/system level and ultimately to behavioral and cognitive levels we experience daily. Perhaps the most fascinating phenomena about learning are related to human behaviors or cognitive capabilities. However, solely gathering data at the behavioral or cognitive level often falls short in uncovering the underlying mechanisms. Fortunately, modern experimental technologies now permit direct observations at the single-molecule or single-cell level. A significant challenge then lies in bridging these granular experimental findings on the lower level to the behavioral and cognitive level. Such efforts would reveal what are the ”relevant” components at molecular or cellular level that are esstential for learning at higher levels.

It is noteworthy that in studying learning phenomena, the mechanistic approach is often employed at the molecular/cellular/circuit level, whereas the normative approach is more

commonly used at the behavioral/cognitive level. Thus, bridging learning across different spatial scales would also reconcile the two viewpoints and offers a coherent picture of how learning happens in the brain.

In the next three chapters, I will present three works in this direction. Along the way, various methods in statistical physics have been found particularly useful, including Langevin dynamics, mean-field theory, replica methods, among others. Given that statistical physics seeks to bridge physical phenomena across scales, the works presented here can be viewed as an extension of the traditional statistical physics framework, applied to one of the most intricate and heterogeneous systems known: the brain.

Below is a summary of the main results:

In chapter 2, we present a quantitative model that links molecular diversity of synapses to its functions on cellular level. These works show that synapses that vary widely across a number of axes (including size, vesicle release rate and calcium sensors) can be captured by a common physical framework, which further led to a quantitative understanding of how diverse forms of synaptic plasticity are constrained by molecular properties of synapses.

In chapter 3, we investigate the effects of plasticity rules measured on single-synapse level on network dynamics. By formulating a mean-field description for spiking neural networks under synaptic plasticity, we show that stability of old memories, while in the presence of acquisition of new memories, requires the plasticity rule to operate on the physiological conditions. The statistical-physical framework developed here can be applicable to a range of other contexts where learning relies on the timing of the relevant events.

In chapter 4, we formulate a recurrent network model for high-dimensional predictive processing. Mathematical analysis of the network model indicates that the most accurate and robust predictive processing in natural conditions arises in a network operating in a regime of loose excitatory/inhibitory balance. This network exhibits a functional desegregation of stimulus and prediction-error representations at the cellular-level. These results illuminate how the perception of one sensory modality (e.g. vision) is shaped by other sensory experiences (e.g.

audition) and offer inspirations for new architectures for unsupervised learning.

Chapter 2

Learning on Molecular/Cellular Level: A Theory of Synaptic Transmission

Abstract

Rapid and precise neuronal communication is enabled through a highly synchronous release of signaling molecules neurotransmitters within just milliseconds of the action potential. Yet neurotransmitter release lacks a theoretical framework that is both phenomenologically accurate and mechanistically realistic. Here, we present an analytic theory of the action-potential-triggered neurotransmitter release at the chemical synapse. The theory is demonstrated to be in detailed quantitative agreement with existing data on a wide variety of synapses from electrophysiological recordings *in vivo* and fluorescence experiments *in vitro*. Despite up to ten orders of magnitude of variation in the release rates among the synapses, the theory reveals that synaptic transmission obeys a simple, universal scaling law, which we confirm through a collapse of the data from strikingly diverse synapses onto a single master curve. This universality is complemented by the ability of the theory to readily extract, through a fit to the data, the kinetic and energetic parameters that uniquely identify each synapse. The theory provides a means to detect cooperativity among the SNARE complexes that mediate vesicle fusion and reveals such cooperativity in several existing data sets. The theory is further applied to establish connections between molecular constituents of synapses and synaptic function. The theory allows competing hypotheses of short-term plasticity to be tested and identifies the regimes where particular mechanisms of synaptic facilitation dominate or, conversely, fail to account for the existing data for paired-pulse ratio. The derived trade-off relation between the transmission rate and fidelity shows how transmission failure can be controlled

by changing the microscopic properties of the vesicle pool and SNARE complexes. The established condition for the maximal synaptic efficacy reveals that no fine tuning is needed for certain synapses to maintain near-optimal transmission. We discuss the limitations of the theory and propose possible routes to extend it. These results provide a quantitative basis for the notion that the molecular-level properties of synapses are crucial determinants of the computational and information-processing functions in synaptic transmission. The text included here is adapted from the published paper [241]. All appendix can be found in its online version.

2.1 Introduction

Neurons communicate across special junctions – synapses – using neurotransmitter molecules as a chemical signal [224]. Release of neurotransmitters into the synaptic gap occurs when neurotransmitter-loaded vesicles fuse with the membrane of the presynaptic (transmitting) neuron in response to calcium influx during an action potential "spike". Synaptic vesicle fusion is remarkably fast and precise: both the duration of fusion and the time between the trigger and fusion initiation are less than a millisecond [109, 224].

The electrical propagation of information along the axon of the presynaptic neuron (the pre-transmission stage) and the response of the postsynaptic neuron to the chemical signal (the post-transmission stage) have been described by theories that capture phenomenology while connecting to microscopic mechanisms [90, 48]. However, neurotransmitter release, which enables the synaptic transmission itself, lacks a theory that is both phenomenologically accurate and microscopically realistic [222]. This void contrasts with detailed experiments, which have revealed the molecular constituents involved. The key to speed and precision of neurotransmitter release is a calcium-triggered conformational transition in SNAREs (soluble N-ethylmaleimide sensitive factor attachment protein receptors) [104, 13, 35]. The free energy released during the conformational transition is harnessed by SNAREs to pull the membranes of the vesicle and the cell together, reducing the high kinetic barriers that otherwise hinder fusion. Fusion culminates

in the release of neurotransmitters from vesicles into the synaptic cleft (Fig.2.1A).

Here, we present a theory of the action-potential-evoked (AP-evoked) synaptic transmission, which quantitatively reproduces a wide range of data from fluorescence experiments *in vitro* [134, 50] and electrophysiological experiments *in vivo* [16, 85, 208, 236, 23, 252, 145, 29, 225, 247, 205, 125, 55, 161, 66]. The theory yields analytic expressions for measurable quantities, which enables a direct fit to the data. Fitting yields parameters that describe the fusion machinery of each synapse: activation barriers and rates of SNARE conformational transitions at any calcium concentration, the size of vesicle pools, and the number of independent SNARE assemblies necessary for fusion. The analytic result of the theory allows to understand, quantitatively, the remarkable temporal precision of neurotransmitter release. Perhaps the most striking result of the theory is that the peak release rate as a function of calcium concentration can be written, with proper normalization, in a universal form so that data on different synapses – with release rates spanning ten orders of magnitude – collapse onto a single curve. The established universality is especially remarkable given that these synapses have been known to exhibit strikingly different properties in synaptic transmission due to distinct Ca^{2+}-sensors [237, 248] as well as different couplings between the SNAREs and their regulatory proteins or calcium channels [108, 238, 221].

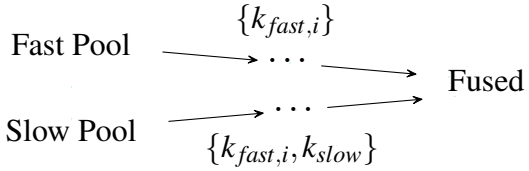
The theory is further applied to relate the properties of neurotransmitter release machinery to the proposed mechanisms of short-term plasticity [196, 99]. A quantitative comparison with experimental data for the paired-pulse ratio enables us to identify the regions where particular mechanisms of synaptic facilitation dominate or, on the contrary, fail to account for the observed facilitation. We establish how the molecular properties of the transmitter release machinery impose constraints on the tradeoff between transmission rate and fidelity, where fidelity measures the ability of a synapse to generate a postsynaptic output in response to a presynaptic input. Finally, we show how the molecular-level properties of synapses determine the optimal synaptic efficacy, or the ability of a synapse to avoid both the transmission errors (lack of a postsynaptic output) and error reads (an output in the absence of an input). Altogether, the theory shows

how the key characteristics of synaptic function – plasticity, fidelity, and efficacy – emerge from molecular mechanisms of neurotransmitter release machinery, and thus provides a mapping from molecular constituents to functions in synaptic transmission.

2.2 Results

2.2.1 Theory

We start from the observation that published data on neurotransmitter release for different synapses and experimental setups [16, 134, 50, 85, 208, 145, 29, 161, 55, 125, 247, 225, 236, 205, 252, 23, 66] can all be encompassed by a unifying kinetic scheme:



(2.1)

In this kinetic scheme, synaptic vesicle fusion proceeds through two parallel reaction pathways. Both pathways contain fast steps of rate constants $\{k_{fast,i}\}$. One of the pathways contains an additional, slow, step of rate constant $k_{slow} \ll \{k_{fast,i}\}$. The pathways originate in the "fast" and "slow" vesicle pools of sizes n_{tot1} and n_{tot2}, respectively. The interpretations of the fast and slow steps as well as the individual states in this unifying kinetic scheme for different experimental setups are summarized in Fig.2.1 and detailed below.

In the context of *in vivo* experiments [85, 208, 145, 29, 161, 55, 125, 247, 225, 236, 205, 252, 23, 66], Eq.(2.1) concretizes into the kinetic scheme in Fig.2.1B and D. The fast pool represents the readily releasable pool (RRP) comprised of vesicles that are docked on the presynaptic terminal (state D) and fuse readily upon Ca^{2+} influx [105]. The slow pool represents the reserve pool (state R), which supplies vesicles to the RRP ($R \rightarrow D$) with slow rate k_2. Fusion of an RRP vesicle ($... \rightarrow F$) requires N independent SNARE assemblies tethering the vesicle at the cell membrane to concurrently undergo a conformational transition. This transition is

Ca^{2+}-dependent and involves a single rate-limiting step [95] of rate constant $k_1([Ca^{2+}])$. Note that N is defined broadly as the critical number of *independent* SNARE assemblies per docked vesicle. Each of the N independent assemblies may consist of a single SNARE or may represent a "super-assembly" of multiple SNAREs that undergo the conformational transition cooperatively [2, 243, 80, 226, 257].

In the context of *in vitro* experiments [134, 50], Eq.(2.1) becomes the kinetic scheme in Fig.2.1C and E. All vesicles are initially docked (states D_1 and D_2) but adopt different morphologies (Fig. 1C) and, consequently, fuse through different pathways [74]. Vesicles in a point contact with the membrane (state D_1) fuse rapidly upon Ca^{2+}-triggered SNARE conformational transition, mimicking RRP vesicles *in vivo*. Vesicles in an extended contact (state D_2) become trapped in a hemifusion diaphragm intermediate (state H), escape from which ($H \to F$) constitutes the slow step k_2.

In all these experiments, the delay due to steps $I_N \to F$ is negligible compared to both fast and slow steps k_1 and k_2. Note that a scheme with N independent and concurrent steps of rates k_1 (Fig.2.1B and C) is equivalent to a scheme with N sequential steps of rates $Nk_1, (N-1)k_1, ..., k_1$ (Fig.2.1D and E).

Despite the differences in the details of the fusion process *in vivo* and *in vitro* described above, the mathematical equivalence of the corresponding kinetic schemes enables their treatment through a unifying theory. We will assume that the calcium influx is triggered by an action potential that arrives at the presynaptic terminal at $t = 0$. The microsecond timescales (much faster than neurotransmitter release) of the opening of voltage-gated Ca^{2+} channels and diffusion of Ca^{2+} ions across the active zone justify treating the $[Ca^{2+}]$ rising as instantaneous. Since the typical width of $[Ca^{2+}]$ profile is $\sim 1 - 10ms$ [20] while most vesicles fuse within $t \sim 100\mu s$ [109], $[Ca^{2+}]$ can be treated as approximately constant during the fusion process. The theory is thus applicable both for step-like and for spike-like $[Ca^{2+}]$ profiles, as well as for responses to long sequences of spikes of the duration shorter than the timescale k_2^{-1} of RRP replenishment. With the above assumptions, the theory is developed in detail in Appendix 1. Below we present

analytic expressions derived from the theory for the key outputs of the experiments that probe synaptic transmission at the single-synapse level *in vivo* and *in vitro*. These expressions relate experimentally measurable characteristics of synaptic transmission to the molecular parameters of synaptic release machinery, thereby enabling the extraction of these parameters through a fit to experimental data.

An informative characteristic of synaptic transmission is the average release rate. Defined as the average (over an ensemble of repeated stimuli) rate of change in the number of fused vesicles, this quantity is the most commonly reported characteristic in experiments on the kinetics of neurotransmitter release [208, 29, 134, 50]. The rate equations for the kinetic scheme in Eq.(2.1) yield the exact solution for the average release rate:

$$\begin{aligned}\frac{d\langle n(t)\rangle}{dt} &= Nk_1n_{tot1}(1-e^{-k_1t})^{N-1}e^{-k_1t} + Nk_1k_2n_{tot2}\sum_{j=0}^{N-1}(-1)^j\binom{N-1}{j}\frac{e^{-k_2t}-e^{-(j+1)k_1t}}{(j+1)k_1-k_2}\\ &\equiv n_{tot1}p_1(t)+n_{tot2}p_2(t),\end{aligned} \tag{2.2}$$

where $p_{1,2}(t)$ are the probability distributions for the fusion time in the fast and slow pathways, N is the necessary number of independent SNARE assemblies, and n_{tot1} and n_{tot2} are the sizes of the fast and slow pools, respectively. We use the standard notation for binomial coefficient $\binom{N}{m} \equiv \frac{N!}{m!(N-m)!}$.

In practice, the average release rate is obtained from the average cumulative release $\langle n(t)\rangle$, which is defined as the average number of vesicles fused by time t and can be measured directly through electrophysiological recording on the postsynaptic neuron [208, 145, 29, 247, 125, 55, 161] or through fluorescence imaging in synthetic single-vesicle systems [134, 50]. Integrating Eq.(2.2) yields the exact solution for average cumulative release:

$$\begin{aligned}\langle n(t)\rangle &= \int_0^t \frac{d\langle n(t)\rangle}{dt}dt = n_{tot1}(1-e^{-k_1t})^N + n_{tot2}\sum_{j=1}^{N}\binom{N}{j}(-1)^{j-1}\left(1-\frac{jk_1e^{-k_2t}-k_2e^{-jk_1t}}{jk_1-k_2}\right)\\ &\equiv n_{tot1}F_1(t)+n_{tot2}F_2(t),\end{aligned} \tag{2.3}$$

where $F_{1,2}(t) = \int_0^t p_{1,2}(t)dt$ are cumulative distributions for the fusion time in the fast and slow pathways can be found in the Appendix 1. *In vivo*, $F_1(t = T) = (1 - e^{-k_1 T})^N$ is the fusion probability for an RRP vesicle after an action potential of duration T [171, 161, 152]. We also derive the full probability distribution of cumulative release by time t (Appendix 1), which, although at present is challenging to measure experimentally, contains more information than the average values in Eqs.(2.2)-(2.3).

Experiments indicate a separation of timescales, $k_2 \ll k_1$ [170, 104], which yields useful asymptotic behaviors for AP-evoked neurotransmitter release. At short times, $t \ll 1/k_1, 1/k_2$, the release rate in Eq.(2.2) is $\frac{d\langle n(t)\rangle}{dt} \sim t^{N-1}$, which can be readily fit to data to extract the number N of independent SNARE assemblies necessary for fusion. At intermediate times, $1/k_1 \ll t \ll 1/k_2$, cumulative release in Eq.(2.3) becomes $\langle n(t)\rangle \approx n_{tot1} + n_{tot2} k_2 t$, which can be used to determine the RRP size, n_{tot1}, by extrapolation [171]. At long times, $t \sim 1/k_2 \gg 1/k_1$, cumulative release is $\langle n(t)\rangle \approx n_{tot1} + n_{tot2}(1 - e^{-k_2 t})$. As expected, the cumulative release on the intermediate and long timescales is independent of the number N of SNARE assemblies and conformational rate k_1 of an assembly as all the fast steps have been completed.

A measure of sensitivity of a synapse to $[Ca^{2+}]$ is the peak release rate [208, 145, 29]. The time at which the peak is reached is found from Eq.(2.2) using $k_2/k_1 \ll 1$: $t_{max} \approx k_1^{-1}\left[\ln N + (n_{tot2}/n_{tot1}\left((N-1)/N^3\right)(k_2/k_1)\right]$. The peak release rate is then

$$\left.\frac{d\langle n(t)\rangle}{dt}\right|_{t=t_{max}} \approx n_{tot1} k_1 \left(1 - \frac{1}{N}\right)^{N-1} \left(1 + \frac{n_{tot2}(N-1)}{n_{tot1} N}\frac{k_2}{k_1}\right). \tag{2.4}$$

Now we must establish an explicit form for the calcium-dependence of the rate constant of SNARE conformational transition $k_1([Ca^{2+}])$ in Eqs.(2.2)-(2.4). We utilize the formalism of reaction kinetics due to Kramers [129] generalized to the presence of a bias field [54]. The formalism treats a conformational transition as thermal escape over a free energy barrier along a reaction coordinate. In the present context, the role of the reaction coordinate is fulfilled by the average number n_{Ca} of Ca^{2+} ions bound to a SNARE assembly at a given $[Ca^{2+}]$, assuming

that this average follows the dynamics of the conformational degree of freedom of the SNARE assembly. The generic shape of the free energy profile with a barrier that separates the two conformational states of a SNARE assembly is captured by a cubic polynomial (Appendix 1 Figure 1). The effect of calcium on the free energy profile is incorporated analogously to the *pH* dependence of the Gibbs free energy of a protein, taking into account both contributions of the electrostatic energy and the entropy [207, 256, 165]. As shown in Appendix 1, the rate constant of the conformational transition of the SNARE assembly is then

$$\begin{aligned} &k_1([Ca^{2+}]) \\ &\quad = k_0 \left(1 - \frac{2}{3}\frac{k_B T n_{Ca}^{\ddagger}}{\Delta G^{\ddagger}} \ln \frac{[Ca^{2+}]}{[Ca^{2+}]_0}\right)^{\frac{1}{2}} \exp\left[\frac{\Delta G^{\ddagger}}{k_B T}\left(1 - \left(1 - \frac{2}{3}\frac{k_B T n_{Ca}^{\ddagger}}{\Delta G^{\ddagger}} \ln \frac{[Ca^{2+}]}{[Ca^{2+}]_0}\right)^{\frac{3}{2}}\right)\right]. \end{aligned} \tag{2.5}$$

Here, k_0 is the rate constant and $\Delta G^{\ddagger}$ is the activation barrier for SNARE conformational transition, and $n_{Ca}^{\ddagger}$ is the number of Ca^{2+} ions bound to a SNARE assembly at the transition state, with all three parameters corresponding to a reference calcium concentration $[Ca^{2+}]_0$. Equation (2.5) provides a quantitative explanation for the remarkable temporal precision of neurotransmitter release. Indeed, the argument of $\exp(\dots)$ is the change in the barrier height at a given $[Ca^{2+}]$ relative to the reference state. The logarithm of calcium concentration, $\ln[Ca^{2+}]$, is the external force that lowers the barrier(concentrations appear logarithmically because the relevant "force" on the molecule comes from the chemical potential, and this helps us to understand how changes in concentration by many orders of magnitude have sensible, graded effects). Equation (2.5) shows that the rate k_1 is exponentially sensitive to this external force, and so are the release rate (Eq.(2.2)) and its peak (Eq.(2.4)) that are both proportional to k_1. This exponentially strong sensitivity of the release rate to the force that drives the release explains, quantitatively, the precisely timed character of synaptic release: synaptic fusion machinery turns on rapidly upon Ca^{2+} influx during the action potential and terminates rapidly upon Ca^{2+}

depletion [223].

Equations (2.2) and (2.5) reveal that the number of independent SNARE assemblies $N = 2$ per vesicle provides the optimal balance between stability and temporal precision of release dynamics [215]. Indeed, at $N = 1$, the release is hypersensitive to sub-millisecond $[Ca^{2+}]$ fluctuations caused by stochastic opening of Ca^{2+} channels (note the high release rate on the sub-millisecond timescale at $N = 1$ in Appendix 1 Figure 2B). On the other hand, at $N > 2$, the peak of release following an action potential is delayed. The optimality of $N = 2$ is further supported by the least squares fit of the experimental data [125] to Eq.(2.3) with different values of N: $N = 2$ results in the smallest fitting errors for all calcium concentrations used in the experiment (Appendix 3 Table 1). However, the theory also reveals that incorporating additional independent SNARE assemblies beyond $N = 2$ may be advantageous for the synapses that require robustness against slower $[Ca^{2+}]$ fluctuations, beyond the sub-millisecond timescale. Indeed, the presynaptic calcium channels are diverse in their intrinsic properties and their interactions with regulatory proteins, and, as the result, generate $[Ca^{2+}]$ fluctuations on a wide range of timescales, $0.5ms - 20ms$ [189, 53]. The shift of the peak release to longer timescales that accompanies an increase in N, as seen in Appendix 1 Figure 2B, allows the synapses to "avoid" correspondingly longer-timescale fluctuations in $[Ca^{2+}]$. This point is illustrated further in Appendix 1 Fig. 2C: in synapses with the larger values of N, the RRP vesicle release (Eq. (2.3)) remains low over longer timescales, thereby providing robustness against slower $[Ca^{2+}]$ fluctuations.

In the presence of cooperative interactions among SNAREs that form super-assemblies, k_1 in Eq. (2.5) represents the effective transition rate of a super-assembly. Appendix 1 Figure 2D illustrates how cooperativity between SNAREs results in a steeper increase of the rate k_1 with increasing $[Ca^{2+}]$, and hence in a faster vesicle release. Specifically, every additional SNARE in the super-assembly is estimated to increase the release rate by a factor of ~ 100 (Appendix 1 Figure 2D), a result consistent with the previous work [154] that utilized a different approach.

Now that we have closed-form expressions for the key characteristics of the neurotransmitter release dynamics in hand, we can establish a universal relation for the sensitivity r of a

synapse to the strength c of the trigger. Nondimensionalization of Eqs. (2.4) and (2.5) gives:

$$r = \exp\left[1-(1-c)^{\frac{3}{2}}\right], \tag{2.6}$$

where $c \equiv \frac{2n_{Ca}^{\ddagger}k_BT}{3\Delta G^{\ddagger}} \ln\frac{[Ca^{2+}]}{[Ca^{2+}]_0}$ and $r \equiv \left(\frac{a}{(1-c)^{1/2}}\frac{d\langle n(t)\rangle}{dt}|_{t_{max}}\right)^{\frac{k_BT}{\Delta G^{\ddagger}}}$ are the dimensionless calcium concentration and peak release rate, and $a \equiv \left(1+\frac{1}{N-1}\right)^{N-1}/(n_{tot1}k_0)$. If the scaling law in Eq.(2.6) indeed captures universal principles of synaptic transmission, data from different synapses should collapse onto the curve given by Eq.(2.6). This prediction is tested in the section "Application of the theory to experimental data" below.

A postsynaptic response to the action potential events is measured by the peak value of the postsynaptic current (PSC). Using the well-established conductance-based model [48], the average of the peak PSC can be shown to be proportional to the total number of released neurotransmitters [109]:

$$\bar{I}_{PSC} = \gamma\langle n(T)\rangle, \tag{2.7}$$

where T is the duration of the action potential ($\sim 1ms$) and γ depends only on the properties of the postsynaptic neuron. As our focus is on the AP-evoked neurotransmitter release in synaptic transmission, γ can be regarded as a constant and postsynaptic receptor saturation can be neglected, so that $\langle n(T)\rangle$ and $\bar{I}_{PSC}$ can be used interchangeably. Note that the presynaptic factors affect the postsynaptic response through $\langle n(t)\rangle$ as described by Eq. (2.3), and include Ca^{2+}-sensitivity of different Ca^{2+} sensors in SNAREs (captured through N, $n_{tot1}k_0$, $n_{Ca}^{\ddagger}$ and $\Delta G^{\ddagger}$) and the sizes of both vesicle pools (n_{tot1} and n_{tot2}). Equations (2.3), (2.5) and (2.7) relate the presynaptic action potential to the postsynaptic current response and thus complete our framework for synaptic transmission. Detailed derivations of Eqs. (2.2)-(2.7) are given in Appendix 1.

To validate the developed analytic theory, we first compare its predictions to data generated through numerical simulations of the kinetic scheme in Eq.(2.1). A simple least squares fit

reliably recovers input parameters of the simulations (Appendix 2 Figure 1 and Figure 2). Next, we test the robustness of the theory by comparing it to modified simulations, in which deviations from the assumptions underlying Eqs. (2.2)-(2.5) are introduced. The modified simulations incorporate (i) the finite-capacity effect of RRP and (ii) heterogeneity of $[Ca^{2+}]$ among different release sites. For deviations within physiological range, the analytic expressions still reliably recover the input parameters (Appendix 2 Figure 3). Details of the simulations are given in Appendix 2.

2.2.2 Application of the theory to experimental data

The availability of analytic expressions for measurable quantities enables direct application of the theory to experimental data. A fit of the peak release rate vs. $[Ca^{2+}]$ with Eqs. (2.4) and (2.5) was performed for a range of synapses to extract a set of parameters $\{\Delta G^{\ddagger}, n^{\ddagger}_{Ca}, k_0\}$ for each synapse. These parameters were then used to rescale the peak release rate and calcium concentration to get the dimensionless quantities r and c that appear in Eq.(2.6). We utilized the experimental data from *in vivo* measurements on (i) the Calyx of Held, a large synapse (diameter $\sim 20\mu m$) in the auditory central nervous system, at different developmental stages [208, 145, 29, 225]; (ii) parallel fiber - molecular layer interneuron (PF-MLI), a small synapse ($\sim 1\mu m$) in the cerebellum [161]; (iii) the photoreceptor synapse [55]; (iv) the inner hair cell [23]; (v) hippocampal mossy fibre [66]; (vi) the cerebellar basket cell [205]; (vii) the retina bipolar cell [85]; (viii) the chromaffin cell [236]; and (ix) insulin-secreting cell [252], as well as (x) two *in vitro* measurements [134, 50]. Figure 2.2 demonstrates that the data from all these synapses collapse on a single curve given by Eq.(2.6), consistent with the prediction of the theory. Even though these synapses have been known to have a huge variation in their release rates (up to ten orders of magnitude) due to the different underlying calcium sensors [41, 116, 101, 124] and different couplings between the SNAREs and their regulatory proteins or calcium channels [108, 238, 221], our theory reveals that all these rates can be brought into a compact, universal form (Eq.(2.6)). The universal collapse is an indication that synaptic

transmission in different synapses is governed by common physical principles and that these principles are captured by the present theory. Variability across synapses on the molecular level is captured through the distinct sets $\{\Delta G^{\ddagger}, n_{Ca}^{\ddagger}, k_0\}$ for each synapse. Notably, the generality of Eq.(2.6) spans beyond the context of synaptic transmission: the same scaling has appeared in another, seemingly unrelated, instance of biological membrane fusion – infection of a cell by an enveloped virus [256].

While a single SNARE can maximally bind $n_{Camax} = 4-5\ Ca^{2+}$ ions [192, 35], the fit of some of the experimental data on the calyx of Held analyzed in Fig. 2.2 produces the transition state values of $n_{Ca}^{\ddagger} > 5$ (Appendix 3 Table 2). This result indicates that each SNARE assembly in these synapses is in fact a super-assembly containing two or more cooperative SNAREs. We further note that, since the number of calcium ions bound to a SNARE at the transition state is generally less than the maximum occupancy for the SNARE, $n_{Ca}^{\ddagger} < n_{Camax}$, the synapses with the values of $n_{Ca}^{\ddagger}$ less than but close to 5 are likely to contain SNARE super-assemblies as well. Interestingly, if we assume that these synapses have the optimal number $N = 2$ of the super-assemblies, and note that the typical rate $k_1 \approx 4ms^{-1}$ at $[Ca^{2+}] = 10\mu M$ would require ~ 3 SNAREs per super-assembly (see Appendix 1 Figure 2D), then the theory estimates that each docked vesicle contains 2 superassemblies $\times$ 3 SNAREs/superassembly $= 6$ SNAREs total. This estimate is consistent with the sixfold symmetric structure recently found using cryoelectron tomography analysis in cultured hippocampal neurons [191].

The utility of the theory as a tool for extracting microscopic parameters of synaptic fusion machinery is further illustrated in Fig. 2.3A-E. A fit of *in vivo* data for cumulative release at different levels of $[Ca^{2+}]$ [247] with Eq.(2.3) extracts the rate of conformational transition of the SNARE assembly, $k_1([Ca^{2+}])$ (Fig.2.3A). A fit of the rate with Eq.(2.5) extracts activation barrier and rate at reference concentration $[Ca^{2+}]_0$ (Fig.2.3B) of the SNARE assembly. Fits of *in vitro* data [134, 50] with Eqs. (2.2) and (2.3) are shown in Fig.2.3C and D. In Fig.2.3C, the content mixing occurrence, defined in [134] as the average release rate normalized by the total number of vesicles, $\frac{d\langle n(t)\rangle}{dt}/(n_{tot1}+n_{tot2})$, is fitted with Eq. (2.2). In Fig.2.3D, the rapid burst

magnitude, defined in [50] as the ratio of the numbers of vesicles fused within the first $1s$ and within $50s$ after calcium trigger, $\langle n(t=1s)\rangle \,/\, \langle n(t=50s)\rangle$, is fitted with Eq. (2.3). Figure 2.3E demonstrates that Eq.(2.5) yields a significantly better agreement with the experimental data on the frog neuromuscular junction [16] than the empirical fourth-power model [16, 99] that was originally used to describe these data. In contrast to the fourth-power model, Eq.(2.5) accounts for the saturation effect in the dose-response curve of a SNARE assembly at high calcium concentrations (see, e.g. the nonlinearity in the rate as a function of calcium concentration on the double logarithmic plots in Fig. 2.2 and Fig. 2.3B).

The parameter values extracted from the fits in Fig. 2.2 and Fig. 2.3 as well as the least-square fitting algorithm for extracting these parameter values are provided in Appendix 3.

2.2.3 Linking molecular mechanisms to synaptic function

Short-term plasticity

Synaptic plasticity, or the ability of synapses to strengthen or weaken over time depending on the history of their activity, underlies learning and memory [196, 12]. A measure of synaptic strength is the peak of the post-synaptic current, which, in turn, is proportional to cumulative release (Eq. (2.7)). The change in synaptic strength that lasts for less than a minute, known as short-term plasticity [196], can be assessed through the paired-pulse ratio, or the ratio of the cumulative release for two consecutive action potentials of width T (typically $T \sim 1/k_1 \ll 1/k_2$) that are separated by interpulse interval τ_{int}. The weakening of a synapse, or short-term depression, is typically caused by the decrease of RRP size due to depletion of vesicles or inactivation of RRP sites [196]. In contrast, the strengthening of synapses, or short-term facilitation, has been attributed to multiple mechanisms [99], including the residual calcium hypothesis put forward in the early studies [109] and recently proposed buffer saturation [120, 169, 27, 11, 110] and Syt7-mediated facilitation [100, 231, 232].

Based on the measured levels of residual calcium concentration of tens to a few hundred nanomolar [260, 166, 100], Eq. (2.5) gives an upper bound of ~ 1.02 for the paired-pulse ratio.

This estimate indicates that the level of residual calcium is far from what is necessary to trigger the large amplitudes of facilitation that are observed in multiple experiments [166, 100], in qualitative agreement with the conclusion in [100].

A more complex version of the residual calcium hypothesis incorporates a facilitation sensor, distinct from the calcium sensor that triggers fusion (usually syt1), which binds to residual Ca^{2+} in between the consecutive action potentials and increases the release probability by interacting with the fusion machinery. Synaptotagmin isoform syt7 has been shown to act as a calcium sensor for facilitation for multiple synapses in the brain [100, 38, 232]. According to the syt7-mediated facilitation scenario proposed in [99], let us assume that syt7 is activated by the residual calcium supplied by the first action potential, and this activation transiently increases the rate of conformational transition $k_1([Ca^{2+}])$ of the main calcium sensor (syt1) by a factor of $\sigma > 1$. Let τ_{res} denote the characteristic timescale on which the new rate $\sigma k_1([Ca^{2+}])$ decays due to the removal of intracellular residual calcium, and let τ_{RRP} denote the recovery timescale of RRP. Assuming the first-order kinetics of calcium removal and RRP recovery, the change in synaptic strength due to the facilitation sensor mechanism can be obtained from Eq. (2.3) as (see Appendix 1)

$$\frac{\langle n_f(T)\rangle}{\langle n_i(T)\rangle} \simeq \left[1 - e^{-\frac{\tau_{int}}{\tau_{RRP}}}\left(1 - e^{-k_1([Ca^{2+}])T}\right)^N\right]\left(\frac{1 - e^{-\left(1+(\sigma-1)e^{-\tau_{int}/\tau_{res}}\right)k_1([Ca^{2+}])T}}{1 - e^{-k_1([Ca^{2+}])T}}\right)^N, \quad (2.8)$$

where the rate constant $k_1([Ca^{2+}])$ is given by Eq. (2.5).

Equation (2.8) enables a quantitative comparison with existing experimental data on a variety of synapses where the activation of syt7 by residual calcium has been proposed as the primary mechanism of facilitation [146, 100, 232]. Figure 2.4 (A-E) shows that the facilitation sensor model in Eq.(2.8) successfully explains, with no additional assumptions, the experimental data on Schaffer collateral, perforant path, corticothalamic, cerebellar granule cell, and retinal ribbon synapses over most of the interstimulus timescales probed in the experiments. At the same

time, the comparison between the data and theory shows that the facilitation sensor mechanism alone fails to explain the data on short ($< 10ms$) timescales for Schaffer collateral and perforant path synapses (Fig.2.4 D, E) as well as on the timescales $> 500ms$ for corticothalamic and granule cell synapses (Fig.2.4 A, B), indicating that other facilitation mechanisms are present and dominate on these timescales. It is worth emphasizing that Eq. (2.8) provides a quantitative model for the syt7-syt1 mechanism, and enables a quantitative test of the facilitation sensor hypothesis, for different synapses through a single, unifying analytic expression. Furthermore, the analytic tractability of the present theory allows the extraction of the parameters that govern the syt7-syt1 mechanism. In particular, the extracted parameters indicate that the syt7-syt1 interaction is strongest ($\sigma = 2.05$) in cerebellar granule cell synapses and weakest ($\sigma = 1.49$) in perforant path synapses. The full list of parameters is included in Appendix 3 Table 3.

In the buffer saturation hypothesis of facilitation [169, 11], Ca^{2+} buffer captures some of the Ca^{2+} ions supplied by the first action potential thereby decreasing the calcium signal for the sensor that triggers fusion. Upon arrival of the second action potential, the fully or partially saturated buffer no longer constrains calcium concentration so that the signal becomes larger, $[Ca^{2+}]_f > [Ca^{2+}]_i$, and can produce facilitation. Let τ_{Ca} denote the characteristic timescale on which the increment in calcium concentration decays due to the dissociation of calcium from the buffer. Assuming the first-order kinetics of the calcium concentration increment and RRP vesicle replenishment, the change in synaptic strength due to the buffer saturation mechanism can be obtained from Eq. (2.3) as (see Appendix 1)

$$\frac{\langle n_f(T)\rangle}{\langle n_i(T)\rangle} \simeq \left[1 - e^{-\frac{\tau_{int}}{\tau_{RRP}}}\left(1 - e^{-k_1([Ca^{2+}]_f)T}\right)^N\right]\left(\frac{1 - e^{-k_1([Ca^{2+}]_f)T}}{1 - e^{-k_1([Ca^{2+}]_i)T}}\right)^N, \tag{2.9}$$

where the rate constant $k_1([Ca^{2+}])$ is given by Eq. (2.5), calcium concentrations during the first and second action potentials are $[Ca^{2+}]_i$ and $[Ca^{2+}]_f = [Ca^{2+}]_i + I_{Ca}e^{-\frac{\tau_{int}}{\tau_{Ca}}}$, and I_{Ca} is the amplitude of the calcium concentration increment due to buffer saturation.

Figure 2.4F shows a quantitative comparison between Eq.(2.9) and the experimental data

on Calyx of Held [166] where buffer saturation has been proposed as the primary mechanism of facilitation [11, 147]. The buffer saturation model in Eq. (2.9) successfully explained, with no assumptions of additional mechanisms, the data over all interstimulus timescales probed in the experiment, thus supporting buffer saturation as the dominant mechanism in mature calyx of Held synapses. Furthermore, the theory enabled the extraction of the dissociation constant for the local calcium buffer and the rate of RRP replenishment from the experimental data (Appendix 3).

The analytic expressions in Eqs.(2.8)-(2.9) can be used to explore, quantitatively, how short-term plasticity is affected by other factors, such as the interplay between the key timescales and the sensitivity of the underlying calcium sensors. For example, Eq.(2.9) predicts that, for fixed interpulse interval τ_{int}, the synapse will exhibit short-term facilitation or short-term depression depending on the ratio of the timescales, τ_{Ca}/τ_{RRP}, as illustrated in Fig. 2.4G [228]. Equation (2.9) further shows that a given synapse may exhibit multiple forms of short-term plasticity when the interpulse interval τ_{int} is varied (Fig. 2.4G). Such coexistence of multiple forms of plasticity has been observed experimentally [196].

A notable feature of Eq. (2.9) is the existence of an optimal value of interpulse interval at which facilitation (at large τ_{Ca}/τ_{RRP}) or depression (at small τ_{Ca}/τ_{RRP}) of synaptic transmission is maximal (Fig.2.4G). Such optimality becomes less pronounced at intermediate values of τ_{Ca}/τ_{RRP} where the synapse exhibits both facilitation and depression (note the curve at $\tau_{Ca}/\tau_{RRP} = 0.4$ in Fig.2.4G), suggesting a more subtle role of short-term plasticity in transmitting transient signals [230, 65].

Equation (2.9) further reveals that a higher Ca^{2+}-sensitivity of the calcium sensor leads to larger facilitation (Appendix 1 Figure 3B), indicating that a high Ca^{2+}-sensitivity of synaptic fusion machinery is essential for the large dynamic range of short-term plasticity. An example of this relationship can be found in [202], and it generally applies to the facilitation synapses where the second spike is associated with higher Ca^{2+} influx, as is the case for the residual Ca^{2+} and buffer saturation mechanisms. Higher $[Ca^{2+}]$ at the second spike causes a larger increase in

rate constant $k_1([Ca^{2+}])$ for a more sensitive synapse compared to the corresponding increase in $k_1([Ca^{2+}])$ for a less sensitive synapse, thus triggering more neurotransmitter release.

Finally, Eq.(2.9) reveals how the molecular-level properties of synapses regulate the facilitation/ depression modes of short-term plasticity (Appendix 1 Figure 3C). The unique properties of neurotransmitter release machinery in different synapses are captured through unique sets of parameters $\{\Delta G^{\ddagger}, n^{\ddagger}_{Ca}, k_0\}$ and τ_{Ca} for each synapse and can reflect different isoforms of synaptotagmin in SNAREs [95, 248], different coupling mechanisms of SNAREs and the scaffolding proteins at release sites [238, 77], or different types of Ca^{2+} buffering proteins present at the presynaptic terminal [213]. These results highlight how the diversity of the molecular machinery for vesicle fusion enables the diverse functions of short-term plasticity [224].

Transmission rate vs. fidelity

An important characteristic of neuronal communication is fidelity of synaptic transmission. Two measures of fidelity can be considered at the single-synapse level for different types of synapses. The probability of spike transmission is a natural measure of fidelity for giant synapses in sensory systems [31] and neuromuscular junctions. The probability of a postsynaptic voltage/current response, beyond the noise level, to a presynaptic spike is a measure of fidelity for small synapses in the central nervous system (CNS) [51]. The probabilistic nature of release mechanisms at synapses is a common origin of synaptic failure [8].

Although the two definitions of fidelity apply to different types of synapses, the present theory allows for a unifying treatment of both phenomena. We assume that the desired postsynaptic response – a postsynaptic spike or a postsynaptic current beyond the noise level – is generated only if the number of released vesicles in response to an action potential exceeds some threshold M. The value of M depends on the density of postsyanptic receptors and the excitability of the postsynaptic neuron [26]. For both types of the postsynaptic response, the probability that the synaptic transmission fails is then obtained from the probability $P\{n(t) = m\}$ that m vesicles

fuse by time t as

$$p_{fail}(T,k_1([Ca^{2+}]),n_{tot1}) = \sum_{m=0}^{M} P\{n(T)=m\} \simeq \sum_{m=0}^{M} \binom{n_{tot1}}{m} F_1(T)^m (1-F_1(T))^{n_{tot1}-m}, \tag{2.10}$$

where $F_1(T) = (1-e^{-k_1([Ca^{2+}])T})^N$. Since the presynaptic neuron cannot generate a second spike during time $[0,T]$, $f \equiv 1/T$ represents the maximum transmission rate. Equation (2.10) predicts that a higher maximum transmission rate f results in a higher probability of transmission failure p_{fail} and thus lower fidelity $(1-p_{fail})$. This trade-off between the maximum rate and fidelity in synaptic transmission is shown in Fig.2.4H. Consistent with intuitive expectation, Eq. (2.10) further predicts that, for a given maximum transmission rate, the probability of transmission failure can be constrained by the RRP size n_{tot1} and/or SNARE conformational rate $k_1([Ca^{2+}])$ (Fig. 2.4H).

Equation (2.10) allows us to make a quantitative statement regarding the molecular-level constraints on the fidelity of synapses of different sizes. Faithful spike transmission implies that the threshold M for postsynaptic response is smaller than the average cumulative release, $M < \langle n(T)\rangle = n_{tot1}F_1(T)$. Then, by the Chernoff bound for Eq. (2.10) [235],

$$p_{fail}(T,k_1([Ca^{2+}]),n_{tot1}) \leq e^{-\alpha n_{tot1}(\frac{F_1(T)}{\alpha}+\ln\frac{\alpha}{F_1(T)}-1)}, \tag{2.11}$$

where $\alpha \equiv M/n_{tot1}$. Because both M and n_{tot1} scale linearly with the area of synaptic junctions [167, 160, 91], it is reasonable to assume that $\alpha = M/n_{tot1} < F_1(T)$ is kept at an approximately constant level for different synapses. Since $F_1(T)/\alpha + \ln(\alpha/F_1(T)) - 1 > 0$, the probability of synaptic failure decreases exponentially as the RRP size n_{tot1} increases. Thus, it follows from Eq. (2.11) that larger synapses tend to be significantly more reliable, i.e., have an exponentially smaller probability to fail, than smaller synapses in transmitting signals [51].

Synaptic efficacy

Equations (2.10)-(2.11) show that synaptic strength can be increased, i.e. failure suppressed, by increasing the RRP size or decreasing the threshold for eliciting postsynaptic response. However, a high synaptic strength increases the probability of an error read, i.e. a postsynaptic response generated without a presynaptic spike. We will now establish the condition for the optimal synaptic strength through the balance of probabilities of failure (no postsynaptic response to an action potential) and error read (postsynaptic response in the absence of an action potential). Let $[Ca^{2+}]_{rest}$ and $[Ca^{2+}]_{AP}$ be the calcium concentrations at rest and during the action potential and q the probability of firing an action potential by the presynaptic neuron. The total probability of transmission error is

$$P(error) = \underbrace{qp_{fail}(T, k_1([Ca^{2+}]_{AP}), n_{tot1})}_{\text{no postsynaptic response after presynaptic spike}} + \underbrace{(1-q)(1-p_{fail}(T, k_1([Ca^{2+}]_{rest}), n_{tot1}))}_{\text{postsynaptic response without presynaptic spike}}. \tag{2.12}$$

Here, we consider the long-term (minutes to days) change in synaptic strength, known as long-term plasticity, through the presynaptic mechanisms and is predominantly due changes in the RRP size, n_{tot1}, which has been shown to be regulated through retrograde signaling according to the threshold M on the postsynaptic side [82, 253, 159, 12]. Synaptic efficacy, $1 - P(error)$, measures the ability of the synapse to faithfully transmit signal. The optimal RRP size is obtained by minimizing the transmission error in Eq.(2.12):

$$n^*_{tot1} = \left\lceil M\left(1 + \frac{\ln \frac{F_{AP}}{F_{rest}}}{\ln \frac{1-F_{rest}}{1-F_{AP}}}\right) + \frac{\ln \frac{q}{1-q}}{\ln \frac{1-F_{rest}}{1-F_{AP}}} \right\rceil, \tag{2.13}$$

where $\lceil x \rceil$ denotes ceiling, i.e. the smallest integer greater than or equal to x, and $F_{AP} = (1 - e^{-k_1([Ca^{2+}]_{AP})T})^N$ and $F_{rest} = (1 - e^{-k_1([Ca^{2+}]_{rest})T})^N$ are the fusion probabilities during the action potential and at rest. Equation (2.13) predicts that, as the synapse is stimulated more

frequently (q increases), a larger RRP size is needed for the optimal performance, i.e. the optimal RRP size and hence the optimal synaptic strength increase, resulting in long-term potentiation on the presynaptic side.

How far can the RRP size deviate from its optimal value without a significant loss of synaptic efficacy? The range of RRP sizes for near-optimal performance can be estimated through the Chernoff bound for Eq. (2.12):

$$P(error) \leq qe^{-\alpha n_{tot1}(\frac{F_{AP}}{\alpha}+\ln\frac{\alpha}{F_{AP}}-1)} + (1-q)e^{-(1-\alpha)n_{tot1}(\frac{1-\alpha}{F_{rest}}+\ln\frac{F_{rest}}{1-\alpha}-1)}. \tag{2.14}$$

According to Eq. (2.14), for synapses that are large ($n_{tot1} \gg 1$) and sufficiently sensitive to Ca^{2+} ($F_{AP}/F_{rest} \gg 1$), the error probability is exponentially small and thus insensitive to changes in the RRP size n_{tot1}. Specifically, the near-optimal range for n_{tot1} can be estimated from $F_{rest} \lesssim \alpha \lesssim F_{AP}$ to be $M/F_{AP} \lesssim n_{tot1} \lesssim M/F_{rest}$. Since $1/F_{AP} \ll 1/F_{rest}$, this range is broad, indicating that large synapses do not need to fine-tune their RRP size in order to maintain near-optimal transmission. This robustness in synaptic transmission is illustrated in Fig. 2.4I.

2.3 Discussion

The capacity of neurons to transmit information through synapses rapidly and precisely is the key to our ability to feel, think, or perform actions. Despite the challenge posed for experimental studies by the ultrashort timescale of synaptic transmission, a number of recent experiments *in vivo* [85, 208, 145, 29, 161, 55, 125, 247, 225, 236, 205, 252, 23, 66] and in reconstituted systems [134, 50] demonstrated the ability to probe the kinetics of synaptic transmission at the single-synapse level. By design, these experiments generate pre-averaged data that encode unprecedented information on the molecular mechanisms of synaptic function, which is lost in the data that are averaged over multiple heterogeneous synaptic inputs. However, decoding this information requires a quantitative framework that would link the quantities that are measured in the experiments to the microscopic parameters of the synaptic release machinery.

Here, we presented a statistical-mechanical theory that establishes these links.

2.3.1 Analytic theory for synaptic transmission

Our theory casts the synaptic fusion scenarios observed in different experimental setups into a unifying kinetic scheme. Each step in this scheme has its mechanistic origin in the context of a given experimental setup. In the context of *in vivo* experiments, distinct vesicle pool dynamics are taken into account [5, 253, 105] to quantitatively explain the different timescales observed in the vesicle release dynamics [104, 172, 201]: vesicles from the readily releasable pool (RRP) fuse readily once the critical number of SNARE complexes undergo conformational transitions upon Ca^{2+} influx (fast step), while the reserve pool supplies vesicles to the RRP (slow step). In the context of *in vitro* experiments, different timescales in vesicle release dynamics are due to the observed distinct states of docked vesicles [50, 133, 74]: the vesicles that are in a point contact with the membrane fuse readily upon Ca^{2+}-triggered SNARE conformational transition (fast step), while the vesicles that are in an extended contact become trapped in a hemifusion diaphragm state prior to fusing with the membrane (slow step). Although the presence of these distinct docked states *in vivo* is still under debate [173, 35], the realization that both of the fusion scenarios can in fact be mapped onto the same kinetic scheme allowed us to capture these scenarios through a unifying analytical theory. The fact that each fusion step in the kinetic scheme has a concrete mechanistic interpretation makes the theory directly predictive in both *in vitro* and *in vivo* experiments.

The calculated measurable quantities include: (i) cumulative release, which quantifies the number of vesicles fused during a given time interval following the action potential, (ii) temporal profile of the release rate, which measures the rate of change in the number of fused vesicles, (iii) peak release rate, which is a measure of sensitivity of a synapse to the trigger, and (iv) the calcium-dependent rate of SNARE conformational change. A least-squares fit of data with these expressions yields the activation energy barrier and rate constant for SNARE conformational change at any calcium concentration of interest, the critical number of SNARE

assemblies necessary for fusion, and the sizes of the readily releasable and reserve vesicle pools.

Since the pioneering efforts to quantitatively describe synaptic transmission [109, 52], multiple models have been developed, such as the "five-site" model and its variants [120, 208, 29, 205, 125, 236, 23] and the dual Ca^{2+} sensor models [225, 185]. These models provided valuable insights into the action-potential-triggered neurotransmitter release in the particular synapses for which they have been developed. However, the existing models have at least two fundamental limitations. First, the system-specific nature of these models limits their applicability beyond specific systems, so that the description of synapses with different calcium-response properties requires the use of different models. In contrast, the present theory is applicable to a wide variety of synaptic types, despite the differences in their fusion pathways, different calcium sensors that they implement [248] and different couplings between their regulatory proteins [108, 77]. Indeed, recent experiments have suggested that the calcium-response properties of synapses are much more diverse than had been thought previously [183, 76, 211]. Second, the existing models did not produce analytic expressions for the key observables that emerge from the experiments, which limits the predictive value of these models, their utility in extracting information from the experiments, and their ability to reveal the organizing principles of synaptic transmission. In contrast, the present theory yields analytic expressions for the key measurable characteristics of synaptic transmission, which can be used as the tools for extracting the essential molecular parameters of synaptic release machinery through a direct fit to experimental data. Thus, the predictive power of the present theory in describing synaptic transmission in vastly different synapses through a unifying framework is complemented by the utility of the theory as a tool for extracting the molecular parameters that uniquely identify each synapse. The theory links the underlying molecular diversity of synapses to the distinct phenomenological responses observed in experiments, and thus constitutes a constructive step toward a yet more complete description of synaptic transmission [222].

The theory presented here has several limitations. (i) Our treatment of the vesicle replenishment rate k_2 as a constant is justified by its weak sensitivity to the intracellular calcium

concentration compared to that of k_1, as found in recent experiments [247, 123, 132]. However, in the response to a tetanic stimulus, where the asynchronous component of the release becomes dominant, the calcium-dependence of k_2 may no longer be negligible. Explicitly taking this dependence into account in the theory will allow the extraction of the parameters for post-tetanic potentiation. (ii) The theory describes synaptic transmission at the level of a single synapse. The theory was motivated by the experimental setups that are capable of probing synaptic transmission at the single-synapse level and is applicable both to giant synapses with many active zones in sensory systems [31] and to small synapses with few active zones in the brain [83] (Fig.2.2). However, a postsynaptic neuron usually receives inputs from many synaptic connections, and the cellular response is an integration of these inputs. The analytic expressions presented above can be directly applied to integrated multiple synaptic inputs in the cases where the molecular features of the presynaptic and postsynaptic sides are similar across the synapses, e.g. when the synapses originate from the same axon and connect to nearby dendritic regions of a postsynaptic neuron [33]. The theory can be extended to account for the effects of heterogeneous presynaptic inputs by applying the derived expressions to each synapse separately with an individual set of microscopic parameters for each synapse. (iii) We treated the postsynaptic response as a linear function of neurotransmitter release (Eq. (2.7)). Such a treatment is sufficient to explain the experimental data on neurotransmitter release (Fig. 2.2 and Fig. 2.3) and the paired-pulse ratio in short-term plasticity (Fig. 2.4) through a single, unifying framework. The theory can be extended to account for the nonlinearity of postsyanptic response by replacing Eq.(2.7) with a relevant nonlinear function. Such an extension will enable the elucidation of the details of active dendritic integration of heterogeneous synaptic inputs.

2.3.2 Ca^{2+}-dependent rate of SNARE conformational transition from Kramers theory

The rate-limiting step in the initiation of fusion of the synaptic vesicles that are docked on the presynaptic membrane is the conformational transition of the critical number of SNARE

assemblies tethering the vesicles to the membrane [104]. We derived the calcium-dependence of the SNARE conformational rate from the classical reaction-rate theory [129] which we generalized to include an external trigger – calcium influx. The resulting analytic expression reveals that the SNARE conformational rate, and hence both the vesicle release rate and the peak of the release rate, are all exponentially sensitive to the force that drives the release – the logarithm of calcium concentration (the logarithmic scale arises naturally due to the several-orders-of-magnitude changes in $[Ca^{2+}]$ following an action potential). This result provides a quantitative explanation for the remarkable synchrony of synaptic vesicle fusion: since the rising of calcium concentration after an action potential occurs on a microsecond timescale and is thus essentially instantaneous on the timescale of synaptic release, the exponential sensitivity of the release rate to this nearly-instantaneous trigger ensures an ultra-rapid initiation of vesicle fusion upon calcium influx. Likewise, the exponential sensitivity of the release rate to the trigger ensures that the fusion process terminates rapidly upon calcium depletion [35].

Unlike the conventional model due to Dodge and Rahamimmoff [52] that postulates $k_{SNARE} \sim [Ca^{2+}]^4$, our expression in Eq.(2.5) naturally accounts for the saturation effect at intermediate-to-high calcium concentrations (Fig. 2.2 and Fig. 2.3B), which is the typical regime for the AP-evoked neurotransmitter release. In the limit of $\ln\left(\frac{[Ca^{2+}]}{[Ca^{2+}]_0}\right) \ll 1$, the asymptotic expansion of Eq.(2.5) recovers the power-law $k_1([Ca^{2+}]) \sim \left(\frac{[Ca^{2+}]}{[Ca^{2+}]_0}\right)^{n^{\ddagger}_{Ca}}$, indicating that a power-law description is only valid for the initial rise of the release rate in response to calcium. Moreover, the power exponent $n^{\ddagger}_{Ca}$ is not a universal number (e.g., 4) but rather depends on the details of the molecular constitutes of the SNARE complexes in a given synapse, such as different calcium sensors from synaptotagmin family [248] and different couplings between the regulatory proteins [108, 221] (Appendix 3 Table 2).

2.3.3 Critical number of SNARE assemblies for vesicle fusion

The theory further reveals how the kinetics of vesicle fusion are affected by the critical number of SNARE assemblies per vesicle. Given the lack of general consensus [224, 215, 234,

35], the theory makes no assumptions about the specific number of SNAREs necessary for fusion, and the number itself can serve as a free parameter when sufficient data is available for a robust fit. Interestingly, however, the theory suggests that $N = 2$ independent SNARE assemblies per vesicle provide the optimal balance between stability and precision of release dynamics. Indeed, on the one hand, in the presence of a single SNARE, the high values and an exponentially-steep temporal dependence of the release rate makes the rate very sensitive to sub-millisecond calcium fluctuations, and thus a very fine tuning of the calcium concentration would be necessary to prevent instability of the fusion process. On the other hand, the values of N greater than 2 lead to longer delays in the peak of the release rate following an action potential, thus reducing the temporal precision of vesicle release. Furthermore, a least-squares fit of the release rate from the experiment [125] with the theory at different values of N reveals that $N = 2$ indeed results in the smallest fitting errors for all calcium concentrations. The generality of this result can be determined as more data on the release dynamics for different synapses becomes available. The theory further suggests that incorporating additional SNARE assemblies beyond $N = 2$ may be advantageous for the synapses that require robustness against slow $[Ca^{2+}]$ fluctuations [164].

The theory can account for cooperativity between SNAREs and can help identify the presence of SNARE super-assemblies [191]. Mathematically, this is due to the formal definition of the parameter N as the number of *independent* reaction steps needed for fusion. Each such step may represent a conformational transition of a single SNARE (in the absence of cooperativity) or of a multi-SNARE super-assembly (i.e., an assembly of cooperative SNAREs). The calcium-dependent release rate $k_1([Ca^{2+}])$ in Eq. (2.5) should be regarded as the transition rate for each independent SNARE unit: if individual SNAREs act independently, k_1 is the transition rate of a single SNARE and N is the number of SNAREs per vesicle; alternatively, if multiple SNAREs undergo conformational change cooperatively, k_1 is the effective transition rate of a super-assembly and N is the number of the super-assemblies per vesicle. The theory allows one to detect the presence of super-assemblies through the values of $n_{Ca}^{\ddagger}$ extracted from the fit: if $n_{Ca}^{\ddagger}$ is larger than the number of Ca^{2+} binding sites for a single SNARE ($n_{Camax} = 5$), it is an

indication that a super-assembly of more than one SNARE is present. Applying this criterion produced evidence for the presence of such super-assemblies in several experimental data sets analyzed in this study. More detailed measurements will be needed to get a more direct estimate of the number of SNAREs in each super-assembly. One approach is to perform single-molecule measurements of the kinetics of a single SNARE under different calcium concentrations, fit the resulting rate $k_1([Ca^{2+}])$ with Eq. (2.5) to extract the value of $n_{Ca}^{\ddagger}$ for the single SNARE, and to compare this value with the value of $n_{Ca}^{\ddagger}$ extracted from a fit with Eq. (2.5) of *i*n vivo data to get an estimate for the number of SNAREs in each super-assembly. The theory suggests that synapses may have more than 2 SNAREs while still having the optimal value of $N = 2$: the SNAREs in these synapses may form $N = 2$ super-assemblies, each comprising more than one SNARE.

2.3.4 Universality vs. specificity in synaptic transmission

The fact that, in all chemical synapses, the delay time from the action potential triggering to vesicle fusion is determined by the conformational transition of preassembled SNARE complexes, and that the conformational transition itself occurs through a single rate-limited step, suggests possible universality in synaptic transmission across different synapses despite their structural and kinetic diversity. Our theory made this intuition precise through a nondimensionalized scaling relationship between the peak release rate and calcium concentration (Eq.(2.6)), which is predicted to hold for all synapses irrespective of their variability on the molecular level. In statistical physics, the significance of universality is that it indicates that the observed phenomenon (here, synaptic transmission) realized in different systems is governed by common physical principles that transcend the details of particular systems.

The universal relation was tested using published experimental data on a variety of synapses, including *in vivo* measurements on the Calyx of Held (an approximately $20\mu m$-diameter synapse in the auditory central nervous system) studied at different developmental stages, parallel fiber-molecular layer interneuron (a $1\mu m$-diameter synapse in the cerebellum),

the photoreceptor synapse, the inner hair cell, the hippocampal mossy fiber, the cerebellar basket cell, the retina bipolar cell, the chromaffin cell, and the insulin-secreting cell, as well as a reconstituted system. Despite more than an order of magnitude difference in the size of these synapses, ten orders of magnitude variation in the dynamic range of synaptic preparations, and a range of calcium concentrations spanning more than three orders of magnitude, the data for the sensitivity of the synapses to the trigger collapsed onto a universal curve, as predicted by the theory. The collapse serves as an evidence that the established scaling of the normalized peak release r with calcium concentration c, $r = \exp\left[1-(1-c)^{3/2}\right]$, is indeed universal across different synapses. At the same time, the unique properties of specific synapses are captured by the theory through the distinct sets of parameters of their molecular machinery: the critical number of SNAREs, their kinetic and energetic characteristics, and the sizes of the vesicle pools. The practical value of the theory as a tool for extracting microscopic parameters of synapses was further illustrated by fitting *in vivo* and *in vitro* data for cumulative release and for the average release rate at different calcium concentrations. Compared to previous work based on phenomenological formulas [126], the mechanistic nature of the present theory allows it to be further tested by independently measuring the microscopic parameters of synaptic fusion machinery $\{\Delta G^{\ddagger}, n_{Ca}^{\ddagger}, k_0\}$ through single-molecule experiments [69, 180] and the postsynaptic response through electrophysiological recording experiments.

2.3.5 From molecular mechanisms to synaptic function

We applied the theory to establish quantitative connections between the molecular constituents of synapses and synaptic function. Previous quantitative analyses of the experimental data on short-term plasticity were based either on the empirical fourth-power model [148] or on custom models that are only applicable to specific calcium sensors [120, 185]. The present theory provides analytic expressions for the paired-pulse ratio (Eqs.(2.8) and (2.9)) that can be directly compared with the existing experimental data on a variety of synapses [166, 100, 232]. As an illustration of the functional implications of the theory, we tested two prevalent hypotheses for the

mechanism of synaptic facilitation: syt7-mediated facilitation and buffer saturation. Our results support the facilitation sensor (Syt7) as the dominant mechanism for short-term facilitation over most of the interstimulus timescales in the Schaffer collateral, perforant path, corticothalamic, cerebellar granule cells, and retinal ribbon synapses, in agreement with [100, 232] but contrary to an earlier study that, in particular, have suggested other mechanisms for facilitation in the retinal ribbon synapse [146]. The theory also identified the regimes where the proposed mechanisms fail to account for the observed facilitation. In particular, the syt7-mediated facilitation cannot explain data at $> 500ms$ for cerebellar granule cell and corticothalamic cell synapses, plausibly due to a dominant effect of buffer saturation in this regime [110, 195]. Likewise, the failure of the syt7 mechanism to explain facilitation in Schaffer collateral and perforant path synapses at $< 10ms$ suggests a significant contribution of the calcium current facilitation in this regime [168]. We limited the discussion of the short-term plasticity to the two mechanisms of synaptic facilitation and to the data on the paired-pulse ratio as illustrative examples, but other mechanisms can be explored in an analogous manner. For example, spike-broadening effects [39] and calcium-dependent vesicle recycling [158] can be incorporated into the theory by introducing variations in T and k_2, respectively.

The theory enabled a quantitative description of how short-term facilitation, depression, or coexistence of multiple forms of plasticity in a given synapse emerge from the interplay between the molecular-scale factors such as the timescales of RRP recovery and buffer dissociation as well as the sensitivity of Ca^{2+}-sensors. In contrast to phenomenological models of short-term plasticity [230, 65, 199], the mechanistic nature of the present theory reveals the connection between temporal filtering of synaptic transmission and calcium-sensitivity of synaptic fusion machinery, and shows how diverse short-term facilitation/depression modes emerge from the diversity of the molecular constituents.

While one intuitively expects that there must be a tradeoff between the maximum transmission rate and fidelity of a synapse, our theory turns this intuition into a quantitative relation (Eq. (2.11)). The trade-off relation shows how transmission failure can be controlled by changing

the microscopic properties of the vesicle pool and SNARE complexes. The relation further shows that the probability of synaptic failure decreases exponentially with increasing the synapse size, which makes large synapses significantly more reliable than small synapses in transmitting signals. Furthermore, the established condition for the maximal synaptic efficacy (Eq.(2.13)) reveals that, for large synapses, the parameter range of near-optimal performance is broad, indicating that no fine tuning is needed for these synapses to maintain near-optimal transmission (Fig. 2.4I). This finding may also be relevant to small synapses: although a small size of their individual RRPs makes them less reliable in transmitting signals individually, trans-synaptic interactions that couple many nearby small synapses may result in a large "effective" RRP [12] and thus enable small synapses to collectively maintain near-optimal transmission without fine-tuning. Altogether, the results of the theory provide a quantitative basis for the notion that the molecular-level properties of synapses are not merely details but are crucial determinants of the computational and information-processing synaptic functions [224]. Limitations of the theory and possible routes to generalize it to other settings are also discussed.

Other biological processes, including infection by enveloped viruses, fertilization, skeletal muscle formation, carcinogenesis, intracellular trafficking, and secretion, have features that are very similar to those in synaptic transmission, despite the bewildering number and structural diversity of the molecular constituents involved [84]. These processes occur through membrane fusion that (i) requires overcoming high energy barriers, (ii) is controlled by proteins that undergo a conformational transition once exposed to a trigger, (iii) is facilitated by the energy released during this transition, which reduces the fusion timescale by orders of magnitude. The theory presented here can be generalized to encompass these processes while engaging with the diversity of specific systems. The mapping from molecular mechanisms to cellular function, provided by the present theory, is a step toward a more complete framework that would bridge mechanisms with function at the multicellular scale (e.g. neuronal circuits and tissues) and further at the scale of an organism.

Acknowledgment

Chapter 2, in full, is a reprint of the material as it appears in Wang and Dudko. ”A theory of synaptic transmission.” Elife 10 (2021): e73585. .The dissertation author was the primary author of this paper.

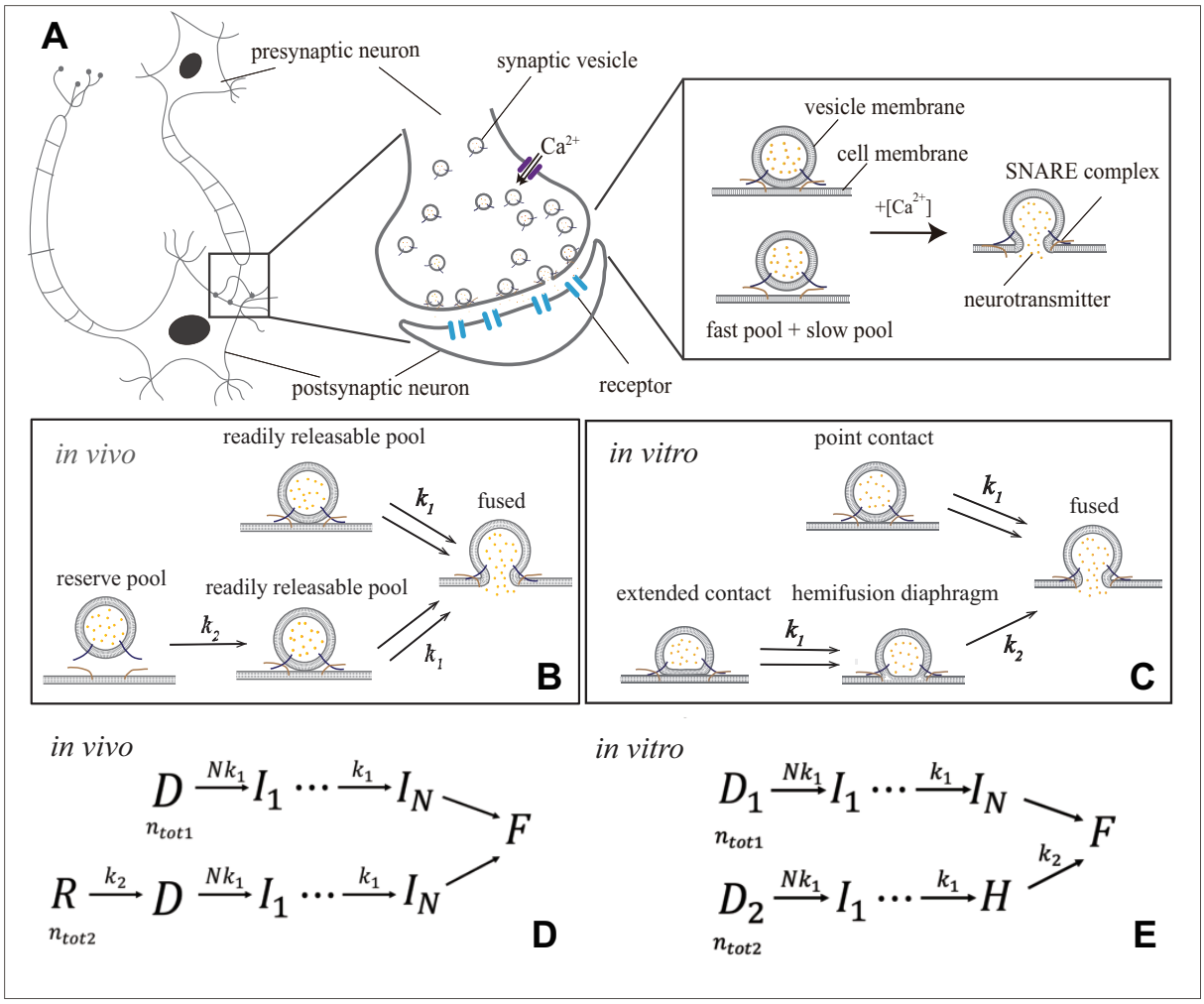


Figure 2.1. Synaptic transmission *in vivo* and *in vitro*. (A) Release of neurotransmitters into the synaptic cleft (diameter $\sim 1-20\mu m$) occurs when neurotransmitter-loaded vesicles (diameter $\sim 30nm$) fuse with the presynaptic cell membrane in response to Ca^{2+} influx during an action potential. Fusion is facilitated by SNARE protein complexes and proceeds via two parallel pathways that originate in the "fast" and "slow" vesicle pools. (B and C) Fusion stages *in vivo* and *in vitro*. SNARE conformational transition constitutes the fast step, k_1. Vesicle transfer from the reserve pool to the readily releasable pool (RRP) *in vivo* and escape from the hemifusion diaphragm *in vitro* constitute the slow step, k_2. (D and E) Reaction schemes for (B) and (C). *In vivo*, state R represents the reserve pool, D the RRP, I_i the state with i independent SNARE assemblies that underwent conformational transitions, F the fused state. *In vitro*, D_1 and D_2 represent docked vesicles with point- and extended-contact morphologies, H the hemifusion diaphragm. Mathematical equivalence of the reaction schemes *in vivo* and *in vitro* enables the treatment through a unifying theory.

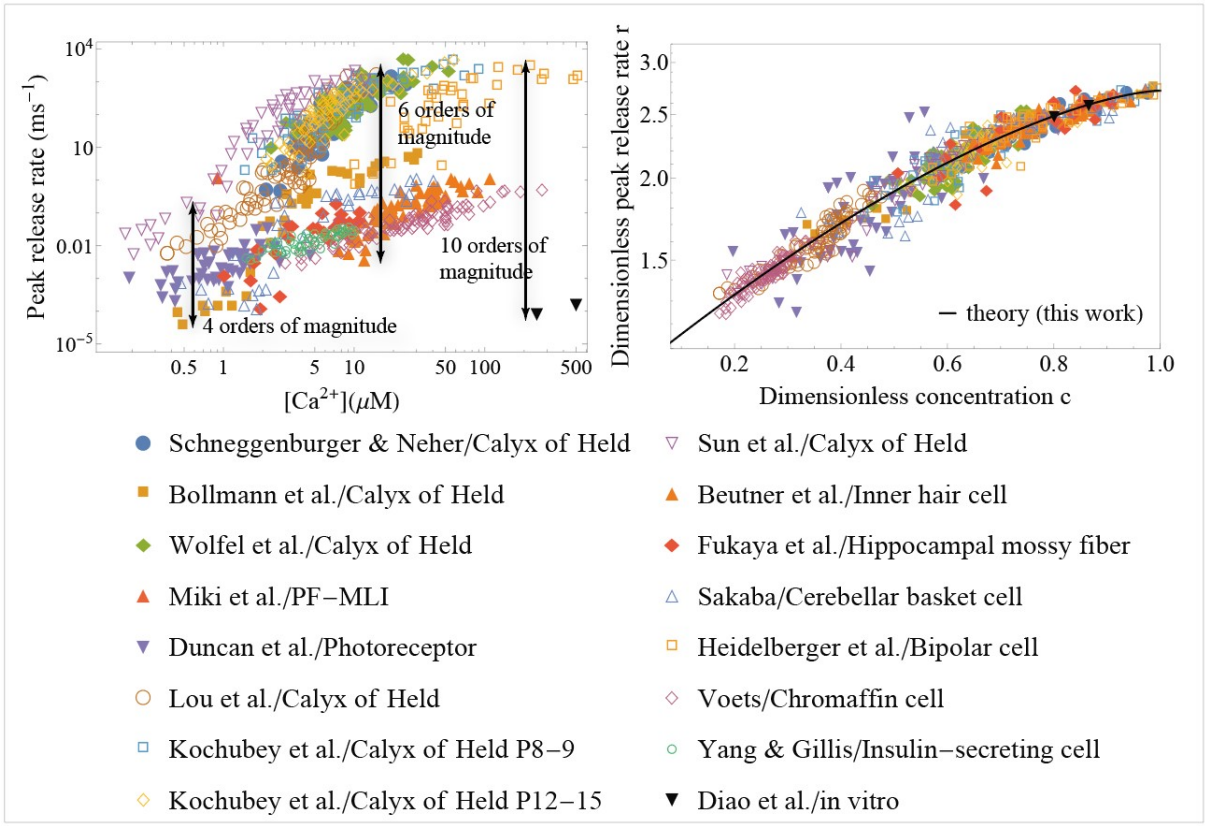


Figure 2.2. Application of the theory to experiments: verifying universality and quantifying specificity. (Left) Measured peak release rate versus calcium concentration for a variety of synapses [208, 29, 161, 55, 50, 125, 247, 145, 85, 23, 66, 205, 236, 252, 225]. (Right) The same data as shown on the left, after the peak release rate and calcium concentration have been rescaled. Despite ten orders of magnitude variation in the dynamic range and more than 3 orders of magnitude variation in calcium concentration (left), the data collapse onto a single master curve, Eq.(2.6) (right). The collapse indicates that the established scaling in Eq.(2.6) is universal across different synapses. The distinct sets of parameters for each of the synapses (Appendix 3 Table 2) demonstrate the predictive power of the theory as a tool for extracting the unique properties of individual synapses from experimental data.

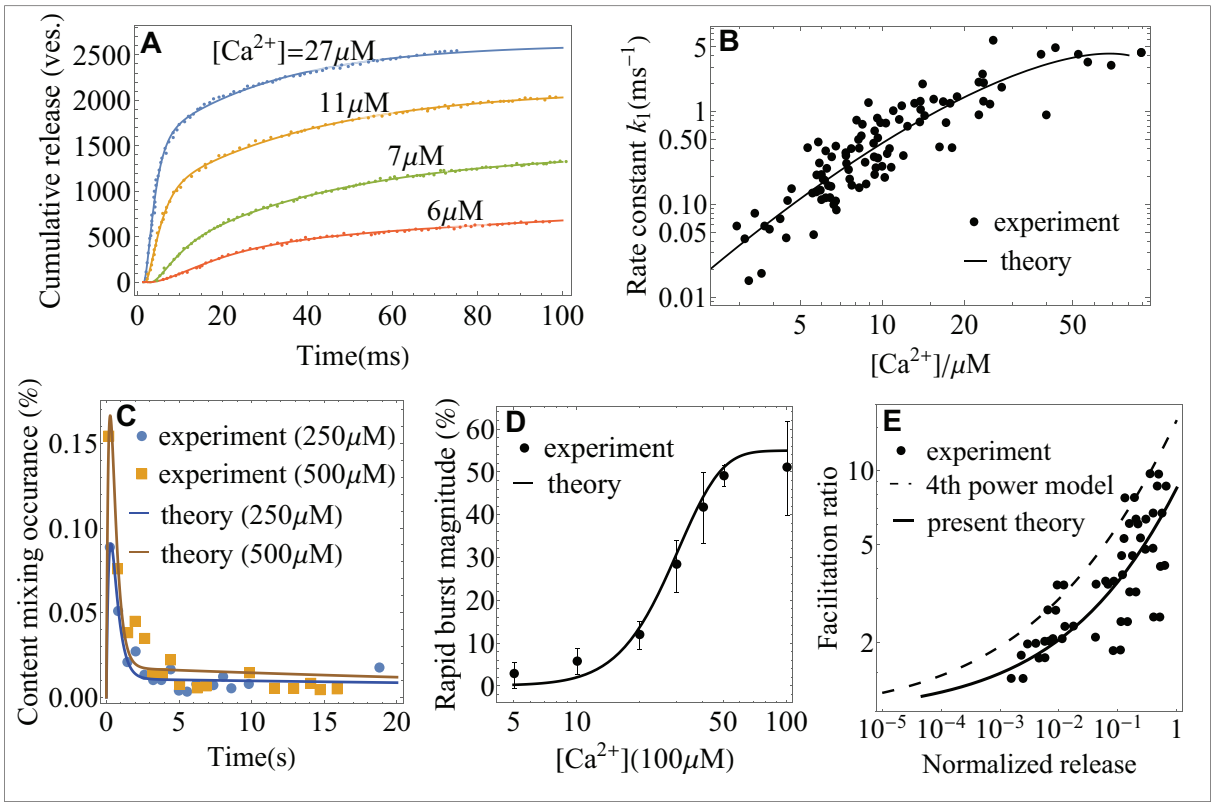


Figure 2.3. Application of the theory to experiments: extracting parameters of synaptic fusion machinery. (A) Cumulative release from *in vivo* experiments [247] on the Calyx of Held (symbols) and a fit with Eq.(2.3) (lines) for different calcium concentrations. (B) Ca^{2+}-dependent rate constant of SNARE conformational transition from *in vivo* experiments [247, 125] and a fit with Eq.(2.5). (C) Content mixing occurrence from *in vitro* experiments [50] and a fit with Eq.(2.2). (D) Rapid burst magnitude from *in vitro* experiments [134] and Eq.(2.3). (E) Facilitation as a function of the ratio of residual and control release (as defined in Appendix 3) from the experiment [16] on the frog neuromuscular junction (symbols) and from the present theory, Eq.(2.5) (solid line). The fourth-power model [16, 99] is also shown for comparison (dashed line). Parameters are shown in Appendix 3.

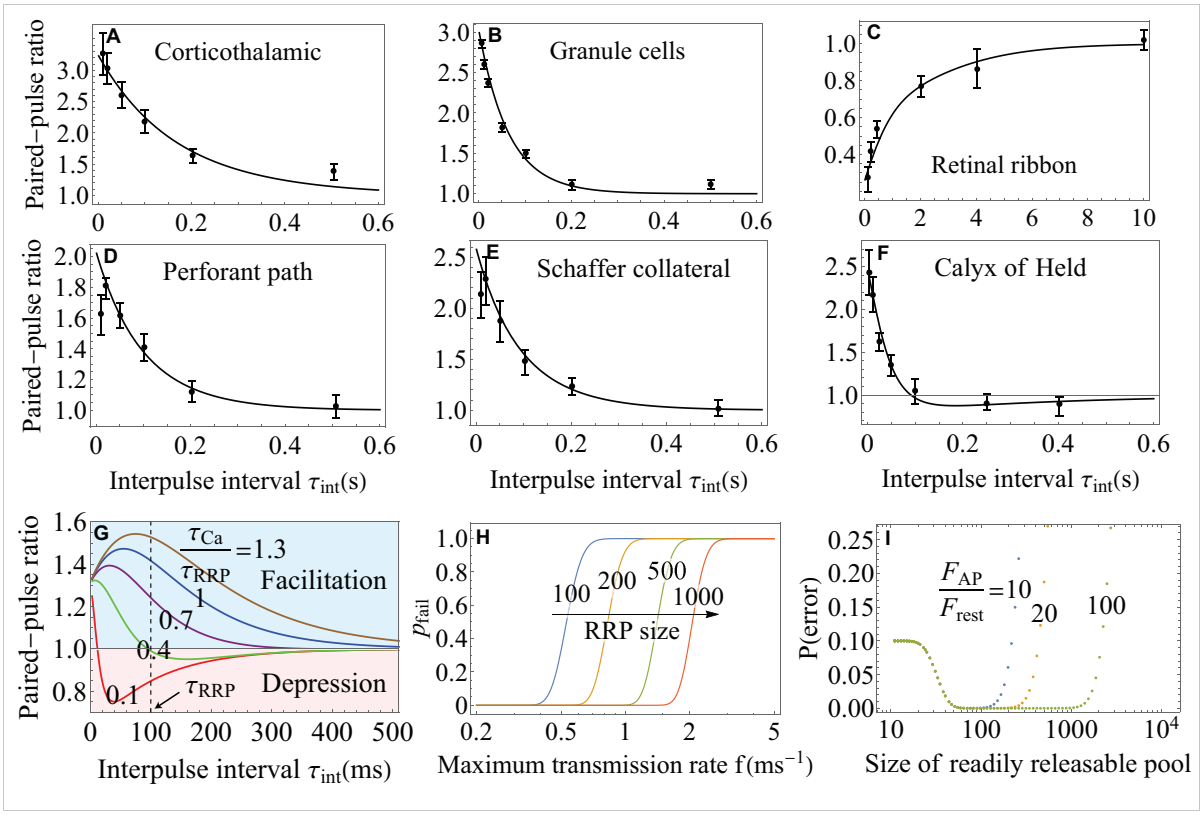


Figure 2.4. Functional implications of the theory. (A-F) The paired-pulse ratio as a measure of short-term plasticity from experiments [100, 232, 166] (symbols) and theory (lines) on a variety of synapses. Equation (2.8) for the Syt7-mediated facilitation captures the data in A-E, and Eq. (2.9) for the buffer saturation mechanisms captures the data in F over most of the interstimulus timescales probed in the experiments. The theory also identifies the regimes where particular mechanisms fail to account for the observed facilitation (A, B, D, E). (G) Paired-pulse ratio predicted by Eq. (2.9). Synapses exhibit short-term facilitation or depression depending on the relative timescales of the recovery of the readily releasable pool, τ_{RRP}, and dissociation of calcium, τ_{Ca}. A given synapse can exhibit multiple forms of short-term plasticity as the time interval τ_{int} is varied. (H) Trade-off between the maximum transmission rate $f = 1/T$ and fidelity $1 - p_{fail}$ from Eq.(2.10) for different RRP sizes. (I) Synaptic efficacy, $1 - P(error)$, from Eq.(2.12). The plateau around the optimal synaptic strength (Eq.(2.13)) indicates that no fine-tuning is required for near-optimal transmission of large synapses. Higher Ca^{2+}-sensitivity F_{AP}/F_{rest} results in broader plateau for near-optimal performance.

Chapter 3

Learning on Network Level via Synaptic Plasticity

Abstract

Fluctuations of synaptic-weights, among many other physical, biological and ecological quantities, are driven by coincident events of two 'parent' processes. We propose a multiplicative shot-noise model that can capture the behaviors of a broad range of such natural phenomena, and analytically derive an approximation that accurately predicts its statistics. We apply our results to study the effects of a multiplicative synaptic plasticity rule that was recently extracted from measurements in physiological conditions. Using mean-field theory analysis and network simulations we investigate how this rule shapes the connectivity and dynamics of recurrent spiking neural networks. The multiplicative plasticity rule is shown to support efficient learning of input stimuli, and gives a stable, unimodal synaptic-weight distribution with a large fraction of strong synapses. The strong synapses remain stable over long times but do not 'run away'. Our results suggest that the multiplicative shot-noise offers a new route to understand the tradeoff between flexibility and stability in neural circuits and other dynamic networks. The text included here is adapted from the published paper [241]. All supplementary materials can be found in its online version.

3.1 Introduction

Many natural processes are triggered by coincidences of two 'parent' events. Examples include firefly flash synchronization [206]; effects of simultaneous environmental stressors [43];

applications of two-photon microscopy [246]; and stimulus-reward associations in reinforcement learning [212, 32]. In neuroscience, co-activation of pre- and postsynaptic neurons plays a crucial role in inducing synaptic plasticity [24, 97], a primary mechanism underlying learning and memory.

The parent processes are often described by event-based models [216], among which the Poisson process is an appealing starting point owing to its memory-less property. Experimental studies show that the aforementioned coincidence-based phenomena often cannot be accurately described as sums over the parent shot-noise (Poisson) processes [43, 97]. Specifically, induction of long-term plasticity was shown to depend strongly on the calcium flux into the postsynaptic neuron [214, 79]. This flux, in turn, depends on coincident spiking activity of pre- and postsynaptic neurons, and is well described by the product of two shot-noise processes [244, 176, 97]. In contrast, most network-level studies of spike-timing dependent plasticity (STDP) typically assume that the synaptic strength change is the sum over contributions of spike-pairs, ignoring cooperative effects between spikes [115, 203, 179, 194, 135]. These models often cannot reproduce realistic spiking activities observed *in vivo* [255].

Motivated by the converging theoretical and experimental evidence, we propose a stochastic process whose fluctuations are triggered by multiplicative interactions between two parent shot-noise processes $[c_\alpha(t),\, \alpha = 1,2]$. The rates of parent events are λ_α, and their amplitudes $a_{\alpha,i}$ are exponentially-distributed with mean A_α. Events are referred to as "spikes", adopting the neuroscience terminology, but they may correspond to events in other domains. Spike-times are denoted $\{t_{\alpha,i}\}$ and may be temporally correlated. We then have,

$$\frac{\mathrm{d}c_\alpha(t)}{\mathrm{d}t} = -\frac{c_\alpha(t)}{\tau_\alpha} + \sum_i a_{\alpha,i}\delta(t - t_{\alpha,i}). \tag{3.1}$$

The decay timescales τ_α define a window during which coincidences can occur.

Our primary interest is a *multiplicative shot-noise* process (denoted $\mathscr{C}(t)$ and henceforth referred to as the *coincidence detector*), whose transient deviations from baseline are driven by

the product $c_1 \times c_2$, with decay timescale $\tau_{\mathscr{C}}$ (Fig. 3.1A),

$$\frac{\mathrm{d}\mathscr{C}(t)}{\mathrm{d}t} = -\frac{\mathscr{C}(t)}{\tau_{\mathscr{C}}} + \eta c_1(t)c_2(t). \tag{3.2}$$

The stochastic calculus of Poisson processes makes it difficult to analyze their products [186]. In contrast, existing nonlinear shot-noise models [227, 155, 204, 59] are equivalent to transformations of a single Poisson process, and are not suitable for studying statistics of coincidences.

We analyze the statistics of the coincidence detector and apply these results to gain insights to a longstanding problem in neuroscience: the stability of recurrent neuronal networks subject to STDP [6]. Here, c_1, c_2 are calcium transients induced by pre- and postsynaptic spikes; and $\mathscr{C}$ is the total calcium flux, which triggers plasticity [97]. Based on spiking network simulations and theoretical analysis of a reduced model, we show that when *individual synapses* in a recurrent network are subject to a nonlinear calcium-based plasticity rule, the empirical *macroscopic* network properties are reproduced (e.g., stable activity patterns, unimodal heavy-tailed synaptic-weight distributions [217]. Further, our results suggest that STDP in itself can support representations that remain stable over timescale of hours, making an important step towards understanding prolonged retention of spatial memories in the face of plasticity and noise [258].

3.2 Statistics of multiplicative shot-noise

For simplicity, we assume that the decay timescales of the parent processes are identical, $\tau_1 = \tau_2 \equiv \tau$. This is consistent with calcium-induced plasticity [177, 97] and cases where the parent processes are generated by similar agents (e.g., firefly flashes [206]).

Solving Eq. (3.1) gives filtered spike-trains (Fig. 3.1A). Using these solutions, we evaluate $\mathscr{C}(t)$ [Eq. (3.2)] at steady-state. When the parent processes are uncorrelated,

$$\mathscr{C} \stackrel{d}{=} \eta\,\tau_{\mathscr{C}} \sum_{i,j} a_{1,i}a_{2,j} \times e^{-\frac{\left|t_{1,i}-t_{2,j}\right|}{\tau}} \times \frac{\mathscr{R}(\min(t_{1,i},t_{2,j}))}{2\delta - 1}. \tag{3.3}$$

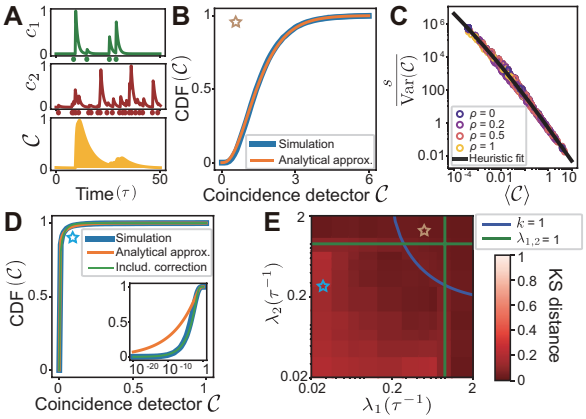


Figure 3.1. Analytical approximation of multiplicative shot-noise statistics. (A) An illustration of the process $\mathscr{C}$, driven by the product $c_1 \times c_2$. (B) The cumulative distribution function (CDF) of $\mathscr{C}$ calculated through method of moments (MM) [Eq. (3.4)] matches simulations well at high firing-rates. (C) Relationship between the log-expectation, mean and variance of $\mathscr{C}$ [Eq. (3.5)]. Data collapses on a line for a range of firing-rates and spike-time correlations (ρ). (D) At low firing-rates, the CDF given by MM matches simulations well for $\mathscr{C} \gtrsim \eta\,\tau_{\mathscr{C}} A_1 A_2$ but poorly for $\mathscr{C} \to 0$. The heuristic MM matches the simulation over the entire range of $\mathscr{C}$. (E) Performance of heuristic MM for uncorrelated spike trains. The errors, measured by KS distance, are ≤ 0.2 for all firing-rates. See [1] for definition of ρ and KS distances for other parameters.

Here, $\delta = \frac{\tau_{\mathscr{C}}}{\tau}$ is the ratio between timescales [Eqs. (3.1, 3.2)], and $\overset{d}{=}$ means "equal in distribution". The natural interpretations of the three factors in Eq. (3.3) are the stochastic amplitudes of synaptic transmission; the temporal window for coincident spikes; and $\mathscr{R}(x) = e^{-\frac{x}{\tau_{\mathscr{C}}}} - e^{-\frac{2x}{\tau}}$ describes firing-rate dependent accumulation of multiple coincidences. Notably, $\mathscr{R}$ represents a departure from summation over spike-pairs [115, 203, 179, 194]. Similar expressions for temporally correlated spike-trains appear in [1].

We begin by formulating an analytical approximation of $P_{\mathscr{C}}$, the distribution of $\mathscr{C}$. At steady-state, the shot-noise process [e.g., c_α in Eq. (3.1)] follows a Gamma distribution [119, 197] with shape and scale parameters $\lambda_\alpha \tau, A_\alpha$. We think of the coincidences of the parent processes as events which drive fluctuations of $\mathscr{C}$. Therefore, we assume $P_{\mathscr{C}}$ can be approximated by a Gamma distribution whose shape (k) and scale (σ) parameters measure the effective rate and amplitude of the coincident spikes.

We used the method of moments (MM), i.e., matching the first and second moments of

$P_{\mathscr{C}}$ to a Gamma distribution, to analytically estimate k, σ. We find for uncorrelated spike-trains,

$$k = \frac{\langle \mathscr{C} \rangle^2}{\mathrm{Var}(\mathscr{C})} = \frac{\lambda_1 \lambda_2 \tau^2}{\frac{\lambda_1+\lambda_2}{\delta+1}\tau + \frac{1}{2\delta+1}},$$
$$\sigma = \frac{\mathrm{Var}(\mathscr{C})}{\langle \mathscr{C} \rangle} = \eta\, \tau_{\mathscr{C}} A_1 A_2 \left(\frac{\lambda_1+\lambda_2}{\delta+1}\tau + \frac{1}{2\delta+1} \right). \tag{3.4}$$

See [1] for expressions including correlations.

We show that MM [Eq. (3.4)] is highly accurate in the high firing-rate regime ($\langle \mathscr{C} \rangle > \eta\, \tau_{\mathscr{C}} A_1 A_2$) by comparing it to numerical simulations (Fig. 3.1E, above blue line). However, for low firing-rates (i.e., $\lambda_1 \tau, \lambda_2 \tau < 1$ which implies $\langle \mathscr{C} \rangle < \eta\, \tau_{\mathscr{C}} A_1 A_2$) and particularly at $\mathscr{C} \to 0$, the MM is inaccurate (Fig. 3.1D). The reason is that for low firing-rates (and $k < 1$), the Gamma probability density diverges at 0. Such a singularity cannot be captured by the mean and variance of $P_{\mathscr{C}}$. Notably, either of λ_1, λ_2 being high suffices for the MM to be accurate, because the high-rate process provides a "background" on top of which the low-rate process can trigger coincidences.

In some applications of our theory, it may be important to accurately estimate $P_{\mathscr{C}}$ at $\mathscr{C} \to 0$, in the low firing-rate regime. We obtain such an estimate by first noticing that in this regime, the maximum-likelihood estimate of the Gamma distribution parameters (k, σ) yields a good approximation of $P_{\mathscr{C}}$. This estimate relies on the log-expectation variable $s = \ln\langle \mathscr{C} \rangle - \langle \ln \mathscr{C} \rangle$, which is indeed sensitive to the singularity at $\mathscr{C} \to 0$. Next we show that the log-expectation, the mean and variance of $\mathscr{C}$ obey a simple relationship, irrespective of the firing-rates and correlation (Fig. 3.1C),

$$\ln\left(\frac{s}{\mathrm{Var}(\mathscr{C})} \right) = -\ln 2 - 2\ln\langle \mathscr{C} \rangle + b(\delta)[-\ln\langle \mathscr{C} \rangle]_+^2, \tag{3.5}$$

where $[x]_+ = \max(x, 0)$ and $b(\delta)$ was fit directly to simulations. We found corrected shape and scale parameters $\tilde{k}$, $\tilde{\sigma}$ by computing s via Eq. (3.5) and using standard maximum-likelihood formulae [1]. We call this the "heuristic MM", and use it when coincidences are rare $k < 1$, and both firing-rates are small $\max(\lambda_1 \tau, \lambda_2 \tau) < 1$. This approximation is accurate for all values of

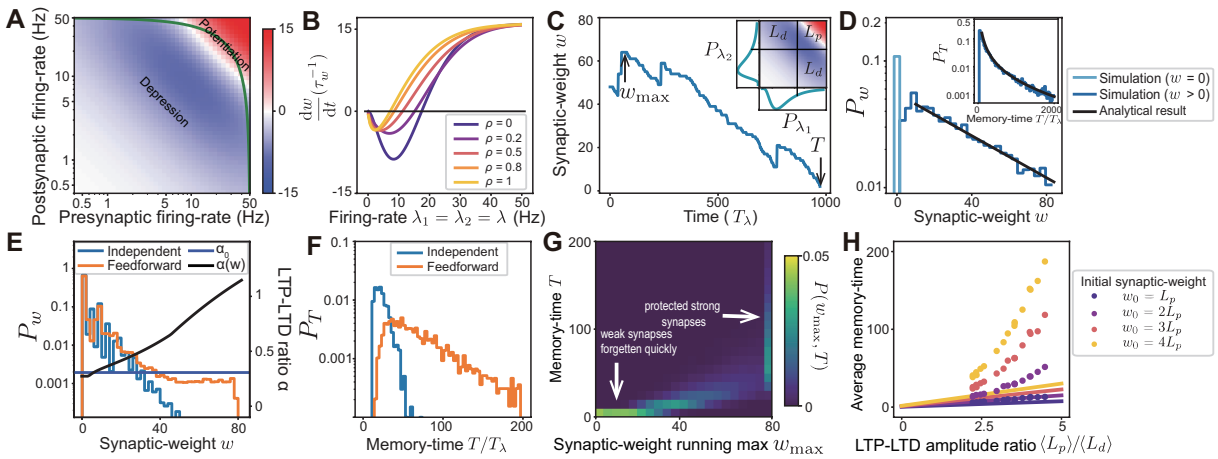


Figure 3.2. Reduced model of synaptic-weight dynamics. Synaptic-weight change [$\frac{dw}{dt}$, Eq. (3.6)] arising from independent (A) and equal (B) pre- and postsynaptic firing-rates. In (A), the temporal correlation $\rho = 0$. The green line indicates $\lambda_1 + \lambda_2 = 50$Hz. (C) Example synaptic-weight trajectory, illustrating the memory-time T and the maximum weight $w_{\max}$. (D) Synaptic-weight distribution, including analytical result for the tail behavior in the weak-synapse scenario. (Inset) Memory-time distribution of a synapse. In the feedforward (strong-synapse) scenario, the synaptic-weight (E) and memory-time (F) distributions have heavier tails. Also shown in (E) is the potentiation-depression ratio $\alpha(w)$. (G) Joint distribution of $(w_{\max}, T)$. Strong synapses are preferentially protected from forgetting. (H) Average memory-time increases linearly with $\langle L_p \rangle / \langle L_d \rangle$ in the independent case, and nonlinearly in the feedforward case. See [1] for parameter values.

$\mathscr{C}$, in the entire parameter space we explored (Fig. 3.1E).

3.3 Network stabilization by multiplicative synaptic plasticity

We now leverage our results to study effects of a multiplicative plasticity on network structure. In this context, $\mathscr{C}$ represents the calcium influx into a neuron, triggered by coincident pre- and post-synaptic spikes. Large influx induces long-term potentiation (LTP; when $\mathscr{C} > \theta_p$), intermediate influx induces long-term depression (LTD; $\theta_p > \mathscr{C} > \theta_d$) [214, 79, 97]. Given the potentiation and depression rates $\gamma_{p,d}$ and thresholds $\theta_{p,d}$, the synaptic-weight dynamics are,

$$\tau_w \frac{dw}{dt} = \gamma_p \Theta(\mathscr{C}(t) - \theta_p) - \gamma_d \Theta(\mathscr{C}(t) - \theta_d) \equiv \Gamma_p - \Gamma_d. \tag{3.6}$$

Note that w has lower bound at 0, and typically $\tau_w \gg \tau_{\mathscr{C}}$. Based on our analytical approximation of $P_{\mathscr{C}}$, we computed the total potentiation/depression rates $\Gamma_{p,d} \simeq \frac{\gamma_{p,d}}{\Gamma(\tilde{k})} \int_{\theta_{p,d}}^{\infty} C^{\tilde{k}-1} e^{-C/\tilde{\sigma}} \mathrm{d}C$, which depend on the spike-train properties and the plasticity rule [Eqs. (3.4, 3.5)].

Neuronal activity *in vivo* undergoes substantial firing-rate fluctuations, generated by external input variability or intrinsic dynamics [40]. Below we formulate a mean-field approximation, reducing the joint dynamics of neurons and synapses and accounting for network structure, to the effective dynamics of a pair of pre- and postsynaptic neurons and the synapse connecting them. The statistical properties of the reduced system recapitulate the network behavior.

In the reduced model, we assume that the neurons' firing-rates (λ_1, λ_2) are sampled from P_λ, and have correlation time $T_\lambda \sim 0.1 - 1$s. P_λ may depend on the synaptic strength, and will be determined self-consistently, accounting for network interactions. During an interval T_λ, the rates (λ_1, λ_2) are approximately constant, so the weight change is $\Delta w = T_\lambda / \tau_w \times (\Gamma_p - \Gamma_d)$. Its distribution $P_{\text{step}}(\Delta w)$ is calculated from P_λ through Eq. (3.6) (Fig. 3.2A,B). Thus, the synaptic dynamics are reduced to a 1D random walk on $w \geq 0$ (Fig. 3.2C) with weight-dependent step-size distribution P_{step}. We identify its steady-state distribution P_w with the synaptic-weight distribution of the network. Importantly, for $T_\lambda \sim 1$s, Δw is not infinitesimal, so the small step-size approximation [115, 203, 149] is invalid. Next we use the mean-field approach to study representative network architectures.

3.3.1 Weak synapses

In this limit, the pre- and postsynaptic firing-rates (λ_1, λ_2) are sampled *independently* of the synaptic-weight w. P_{step} is further assumed to be discrete, such that a synapse can be unchanged/potentiated/depressed by fixed amounts $\Delta w = 0, L_p, -L_d$ with probabilities $\alpha_0, \alpha_p, \alpha_d$. Equivalently, $P_{\text{step}}(\Delta w) = \alpha_0 \delta(\Delta w) + \alpha_d \delta(\Delta w + L_d) + \alpha_p \delta(\Delta w - L_p)$. Such a scenario would be expected in a network switching between "high" and "low" states. For illustration, we assume that L_p/L_d is an integer (see [1] for more general analysis). We define the potentiation-depression ratio $\alpha \equiv \alpha_p L_p / \alpha_d L_d$. When depression dominates $(\alpha < 1)$, using results for random walks

[61, 63], P_w is unimodal, and its tail follows $P_w \sim h(\beta)e^{-\frac{\beta w}{L_d}}$, $w \gg L_p$. The factors $\beta, h(\beta)$ were determined by analyzing the moment generating function of w [1].

We find that the tail becomes heavier as overall potentiation and depression are more closely balanced ($\alpha_d L_d \gtrsim \alpha_p L_p$); or as L_p/L_d is larger with fixed α. Generally, when weight changes (Δw) are non-negligible relative to the mean, the distribution P_w is unimodal, and its tail behavior is sensitive to the high-order statistics of the step-size distribution P_{step}, in contrast to the case of infinitesimal Δw [203].

To study the process of forgetting in the reduced model, we envision a potentiated synapse with initial weight w_0 representing a certain memory. The memory-time $T(w_0)$ is that synapse's first-passage time to 0. Analysis of the random walk statistics [1] gives the average and tail behavior of $T(w_0)$,

$$\langle T(w_0)\rangle = \frac{w_0}{\alpha_d L_d - \alpha_p L_p} \quad \text{and} \quad P_T(t) \sim w_0 t^{-\frac{3}{2}} e^{-\frac{t}{\kappa}}. \tag{3.7}$$

See [1] for expressions of κ. Similarly to P_w, the average memory-time becomes longer and the tail becomes heavier as L_p/L_d increases, with fixed α (Fig. 3.2E).

3.3.2 Strong synapses

When synapses are strong [15, 3], P_λ becomes weight-dependent. The postsynaptic neuron receives *feedforward* weighted presynaptic input (firing-rate λ_1, weight w) and background input from the rest of the network. Both inputs switch between high and low firing-rates. Here, using the heuristic MM to compute $P_{\text{step}}(\Delta w)$ requires knowing how spike-time correlations depend on w and the background input. This relationship was determined by matching the postsynaptic neuron with a leaky-integrate-and-fire neuron driven by presynaptic shot-noise and background Gaussian noise [1]. We then numerically evaluated P_w, P_T at steady-state, showing a substantially heavier tail when compared to the independent case (while fixing the overall potentiation-depression ratio $\langle \alpha(w)\rangle_{P_w}$, Fig. 3.2E,F).

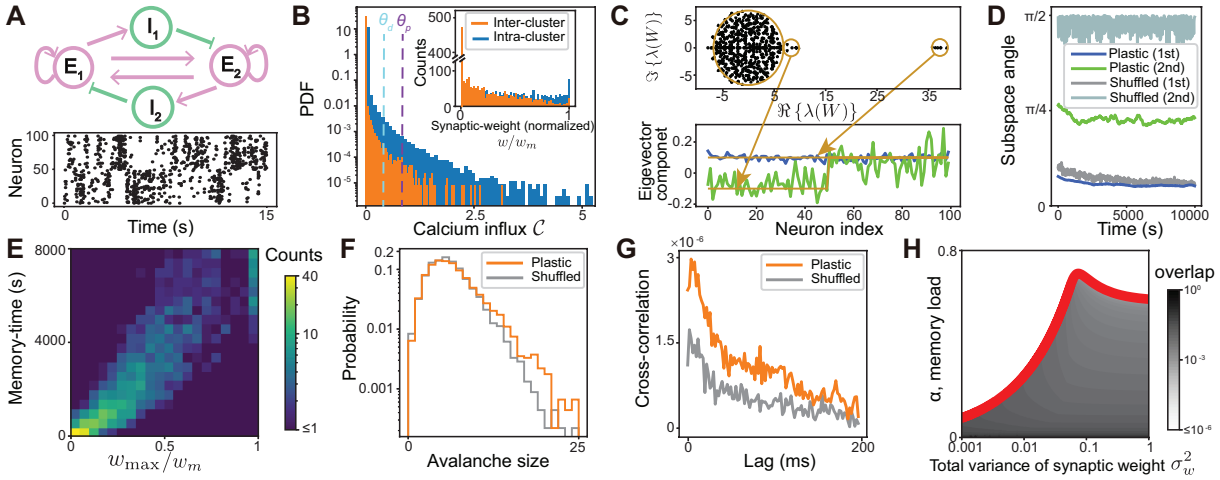


Figure 3.3. Recurrent spiking neural network simulations with synapses subject to the multiplicative plasticity rule. (A) Initial network structure, and steady-state spiking activity. (B) Calcium influx and synaptic-weight (inset) distributions at steady-state, for inter- and intra-cluster connections. (C) Eigenvalues (top) and eigenvectors (bottom) of the $E \to E$ steady-state synaptic-weight matrix. The second outlier suggests that the network's cluster structure is preserved at steady-state. (D) The first two subspace angles between the plastic network and a network with perfect cluster structure as a function of time. The second angle in the plastic network remains far from $\pi/2$ compared to the shuffled network. (E) Joint distribution of $(w_{\max}, T)$ for intra-cluster $E \to E$ connections at steady-state. Similarly to the reduced model (Fig. 3.2C), strong synapses are preferentially protected. (F,G) The plastic network exhibits larger avalanches and stronger intra-cluster spike-timing correlations, compared to the shuffled network. (H) Memory capacity of a Hopfield-like network versus synaptic-weight variability. Synaptic-weights are normalized such that the variance is independent of the memory load α. See [1] for simulation and calculation details.

Inspection of the joint distribution of the synaptic-weight running maximum and the memory-time $(w_{\max}, T)$ suggests that strong synapses are specifically resistant to forgetting (Fig. 3.2G). Moreover, the average memory-time increases nonlinearly with the LTP-LTD amplitude ratio (Fig. 3.2H), compared to an approximately linear increase in the independent case [Eq. (3.7), again matching $\langle \alpha(w) \rangle_{P_w}$]. Similar results were observed in a reduced clustered recurrent network model [1].

Taken together, in the regime where a small number of inputs is sufficient to trigger a postsynaptic response, the multiplicative plasticity rule supports a unimodal synaptic-weight distribution in which strong synapses are preferentially protected from turnover.

3.4 Spiking network simulations

We tested our results by simulating a network of leaky-integrate-and-fire neurons. The network consists of two excitatory (E) clusters which mutually inhibit each other indirectly via inhibitory clusters (Fig. 3.3A, see [1] for networks with > 2 clusters). Initially, intra-cluster $E \to E$ connections are strong while inter-cluster connections are weak. Crucially, the probability that an $E \to E$ connection exists is independent of the cluster assignment. The initial structure may represent two mutually exclusive memories stored in the network that spontaneously switch on timescale of ~ 0.2s (Fig. 3.3A). Structured inhibition is consistent with experiments showing inhibitory stimulus-specific ensembles, and may arise from inhibitory plasticity [117, 71, 62]. We investigated memory retention when $E \to E$ synapses undergo multiplicative plasticity, by examining the steady-state statistics of $\mathscr{C}$, network structure, and dynamics.

In this network, potentiation is more likely in intra-cluster relative to inter-cluster synapses, so the tail of P_w for intra-cluster synapses is heavier (Fig. 3.3B). Yet, notably, P_w is unimodal with only minimal saturation to the upper bound. To examine stability of network structure, we plotted the steady-state $E \to E$ weight matrix spectrum (Fig. 3.3C). The spectral distribution's bulk follows the circular law for a network with independent, random weights, sampled from cluster-specific distributions [7, 1]. Additionally, there are two outlying eigenvalues. The fact that the larger eigenvalue (corresponding to the "DC" eigenvector) does not saturate to its maximum possible value, together with the stability of the switching dynamics, suggests that there is no runaway potentiation of either cluster. The smaller eigenvalue corresponds to an eigenvector that follows from the clustered connectivity. Angles between the plastic network's eigenvectors and those of a network with perfect cluster structure are stable and much smaller than angles computed for a network with shuffled connections, indicating that network structure is preserved despite ongoing plasticity. As predicted by the mean-field analysis, strong synapses are protected from rapid turnover. The dynamics of the plastic network also retain the cluster properties, exhibiting larger intra-cluster spike-time correlations and larger avalanches, than a

shuffled network. Intriguingly, avalanche statistics are closely related to the synaptic-weight distribution [131].

To understand the implications of the multiplicative rule beyond stability, we extended the results in [187] and analytically computed the memory capacity of Hopfield-like network, defined such that the variance of the synaptic-weight distribution is independent of the memory load [1]. Fig. 3H shows that a heavier tail of the distribution, similarly to P_w in Fig. 2E, leads to a marked increase in capacity. Furthermore, we demonstrate in [1] that the multiplicative plasticity rule supports efficent learning of structured connectivity (akin to Fig. 3A), reflecting the structure of an external input.

In [1] we explored the sensitivity of the spiking-network stability results to changes of two key parameters, the potentiation/depression thresholds [$\theta_{p,d}$ in Eq. (3.6)] and the structure of inhibition.

We additionally showed that networks with intrinsically bistable synapses [79] are also stable, but do not exhibit realistic synaptic-weight distributions or activity-dependent protection of strong synapses. Highlighting the importance of the multiplicative rule's statistics, we found that an additive plasticity rule with $\mathscr{C} = c_1 + c_2$ [instead of Eq. (3.2)] rapidly leads to instability, and is unable to efficiently learn the structure of an external input [1].

Our analysis offers insights to the two-timescale problem, where synaptic interactions determine network dynamics on short timescales, and undergo neural activity-dependent modifications on longer timescales. Importantly, we analyze the network in a regime where strict separation of timescales does not hold. Previous studies utilizing plasticity rules where modifications depend (possibly nonlinearly) on sums over pre- and postsynaptic activity, typically resulted in unrealistic synaptic-weight or firing-rate distributions, or required fast homeostatic mechanisms for stability [115, 203, 142, 89, 179, 194, 4, 127]. The multiplicative structure of the plasticity rule analyzed here effectively eliminates modifications due to 'spurious' activity, while specific patterns of activity are responsible for potentiation and learning. The general structure of the multiplicative process introduced here suggests that our results could be applied

to understand nonlinear and adaptive interacting systems in a broad range of scientific fields.

Acknowledgment

Chapter 3, in full, is a reprint of the material as it appears in Wang and Aljadeff. ”Multiplicative Shot-Noise: A New Route to Stability of Plastic Networks.” Physical Review Letters 129.6 (2022): 068101. The dissertation author was the primary author of this paper.

Chapter 4

Learning on Cognitive Level: Desegregation of Neural Predictive Processing

Abstract

Neural circuits across animal species and brain regions are thought to construct internal 'world-models' to guide behavior. The predictive processing framework posits that this is done by generating predictions for sensory signals, and concurrently computing prediction-errors. Extending this framework to complex multi-sensory signals encountered in naturalistic environments is a major challenge in neuroscience. We address this challenge here by investigating how multi-modal and high-dimensional predictive representations can emerge in recurrent neural networks during learning. Our analysis of neural representations formed in the network indicates that the most accurate and robust predictive processing in natural conditions arises in a network operating in a regime of loose excitatory/inhibitory balance. This network exhibits a functional desegregation of stimulus and prediction-error representations, at the cellular-level. We confirmed our model predictions based on recent experiments that probe predictive coding circuits using a more rich set of stimulus associations than done previously. Overall, our study suggests that, in natural conditions, neural representations of internal models are highly distributed, yet structured so as to allow effective readout of behaviorally relevant information. Furthermore, owing to the generality of our model, our study will help decipher how internal models are formed across brain regions, including outside of the neocortex, by linking circuit mechanisms and architectures to functional properties of neural predictive processing. The text here is adapted from the unpublished manuscript. The supplementary materials will be available once the manuscript is online.

4.1 Introduction

Predictive coding, the process of computing the expected values of sensory, motor, and task-related quantities, is thought to be a fundamental operation of the brain [193, 114]. Violation of internally-generated expectations, known as *prediction errors*, is an important neural signal that can be used to guide learning and synaptic plasticity [190, 245]. Signatures of predictive coding, including neural correlates of prediction errors, were identified in multiple brain circuits, and across animal species [57, 113, 114, 239]. Two well-studied examples are motor-auditory predictions [174, 209, 210, 10, 9] and visual-auditory predictions [98, 96, 70] in the mouse cortex. Previous work has proposed that a *canonical cortical microcircuit* underlies the computation of predictions and prediction errors, based on imaging, electrophysiological and genetic experimental tools [174, 209, 114, 184]. While some predictions of this proposed microcircuit were tested and confirmed in restricted scenarios, the hypothesis that the circuit-motif within the mouse cortex is a general mechanism for predictive processing faces a number of challenges:

- **Predictive processing of multi-modal and high-dimensional stimuli:** Typical experimental paradigms to study predictive coding are based on training animals to make a single association between sensory stimuli from different modalities [70], or a single sensory-motor pairing [102, 10]. However, in natural settings, predictable stimuli are typically high-dimensional and multi-modal (e.g., speech production, [92]). In addition, associations between sensory stimuli and different motor commands are often context-dependent. For example, the visual scene may provide a contextual signal that alters the sounds expected from self-motion. While some work in this field has proposed a conceptual framework for computing multi-modal predictions [114, 70], how these computations are learned in specific circuit structures is not understood.

- **Predictive processing in subcortical regions and in brain circuits in invertebrates:** Placebo analgesia can be regarded as *prediction-based* suppression of pain, and has been

suggested to rely in large part on predictive processing of pain information in the periaqueductal gray and rostral ventromedial medulla regions [37]. Here, pain-inducing signals and pain-relief predictions typically consist of sensory inputs from multiple modalities. Other examples of predictive processing outside of the vertebrate brain include cephalopod camouflage behavior, where the displayed high-dimensional skin pattern can be thought as a prediction of the pattern which is best suited to disguise the animal from predators, given the visual scene (which can also be high-dimensional) [249], and more 'classic' predictive computations on motor-visual associations in insects [118]. It is not known whether the neural circuits implementing those computations use analogous strategies to those employed by the mammalian cortex, or altogether different strategies.

- **Predictive processing on multiple timescales:** In some experimental paradigms, days of training are required to observe neural signatures of predictive coding [151, 70]. Those changes to neural responses likely require long-term synaptic plasticity. However, neural responses that are strongly correlated with prediction errors can emerge on much shorter timescales. In the 'odd-ball' paradigm, *stimulus specific adaptation* (SSA) is seen when stimuli become predictable based on statistical regularities of the input sequence [233, 86]. The fact that input sequences lasting tens to hundreds of seconds are sufficient to elicit SSA suggests that reorganization of circuit structure and long-term synaptic plasticity cannot account for these effects [254]. In other scenarios, the brain-body axis was shown to compute predictions of future hunger and thirst states. The intake of food and liquids can cause an almost immediate feeling of satiety, although the absorption process of nutrients and water may take minutes to hours [144, 143]. It is currently not well understood what prediction error responses require circuit reorganization, and what responses rely on other mechanisms that are typically not included in models of predictive coding circuits.

The evidence that computing predictions is an integral part of sensory processing garnered attention from the theoretical neuroscience community. A number of studies proposed recurrent

network models that may perform those computations [28, 46, 103, 88, 87, 162, 163]. Those modeling studies focused on predictions within a single sensory modality, and on predictions of a small number of stimuli. The model proposed so far cannot capture the breadth of phenomena that can be understood within the predictive coding framework across animal species, sensory modalities and brain regions.

A major current gap from both experimental and modeling perspectives is predictive processing in high-dimensions: (*i*) What are the neural representations of predictable and unpredictable sensory variables in natural conditions with rich stimulus ensembles and complex inter-dependencies between stimuli [47, 239]? (*ii*) What are the circuit mechanisms underlying the computation of those representations, and how are they learned? We fill this gap here, specifically focusing on the question of functional segregation of predictive coding [114, 184], and on how predictive coding circuits are formed during learning.

We address these questions by developing a mathematical model to examine the predictive representations in recurrent networks for multi-modal and high-dimensional inputs during and after learning, and by relating this model to experimental data. By generalizing previous models [28, 103, 163] to apply to nonlinear computation of high-dimensional predictions, our study advances the understanding of how neural networks perform predictive processing in scenarios where animals are trained to perform more naturalistic tasks. From a mechanistic perspective, we provide a surprising prediction for the degree of excitation/inhibition balance in the high-dimensional case, and novel insight into the role E/I balance plays in canceling interference between multiple learned stimuli. From a functional perspective, our normative approach suggests that predictive processing of high-dimensional stimuli is robust when the representations of stimuli and of prediction errors are *desegregated* at the cellular-level, and distributed among excitatory and inhibitory neurons.

Overall, our mathematical results and data analysis suggest together that high-dimensional predictive processing in neural circuits is distributed. Since the model proposed here is more general than previous models of cortical predictive processing, and since we provide evidence

for its applicability based on experimental data, we believe that our work reveals principles of predictive processing across species and brain-regions.

4.2 Results

4.2.1 Recurrent networks that learn to generate high-dimensional predictions

We studied the neural representations formed in recurrent neural networks that perform predictive processing of multi-modal sensory and motor inputs. We focus on a typical associative training scenario where animals are presented with pairs of sensory stimuli simultaneously [209, 210, 10] or after a short delay [70]. The stimuli comprising each pair are typically of different sensory modalities (e.g., auditory-visual [70]), or involve a sensory-motor association (e.g., locomotion-auditory [10]). In this scenario, predictive computations are thought to be learned over time through synaptic-weight updates [112, 209, 210, 102, 70, 10]. Our network model consists of N recurrently connected neurons whose firing rates depend nonlinearly on the input current driving their responses (Fig. 1a). The presentation of stimuli to the network is determined by the variables x and y. The strength of the input to each neuron corresponds to the components of the stimulus-specific feedforward synaptic weight vectors $\boldsymbol{w}$ and $\boldsymbol{v}$. There are P stimulus pairs, and when P is of the same order as the number of neurons N, the network is said to perform *high-dimensional* processing.

Before training, the feedforward weight vectors corresponding to each stimulus-pair are random and uncorrelated within the pair (i.e., $\boldsymbol{w} \cdot \boldsymbol{v} = 0$). During training, those weights become correlated ($\boldsymbol{w} \cdot \boldsymbol{v} = \mu$), consistent with experimental evidence of learning-induced functional reorganization of excitatory synaptic connections [122, 42, 56]. Our goal is to define recurrent connectivity that will allow (*i*) reading out the identity of the presented stimulus, (*ii*) predicting the 'missing' stimulus (i.e., predicting y based on x), and (*iii*) evaluating the prediction error (Fig. 1a). Specifically, the weights of the recurrent connections are chosen so as to minimize the

error between the internally generated prediction and the actual stimulus, as well as the overall encoding efficiency (Methods).

We studied neural responses during learning in the *match* condition (where $x = y$), and in the *mismatch* condition (where $x \neq y$). When we compared model neural responses to experimental data, x and y are binary variables corresponding to the presence ($x, y = 1$) or absence ($x, y = 0$) of visual and auditory stimuli [70], visual stimuli and a motor command [102], or auditory stimuli and a motor command [10, 9]. We extended the mathematical formalism to scenarios where more than two stimuli are predictive of each other, and to scenarios where the inputs to the network vary continuously ($0 \leq x, y \leq 1$, e.g., running speed or visual-flow speed, [112, 102], Methods).

Before associative learning occurs ($\mu = 0$), most of the neurons in the network have comparable responses in the match and mismatch conditions. After training ($\mu = 0.9$), match responses are suppressed while mismatch responses are amplified (Fig. 1b). Correspondingly, the ratio of the average firing-rates in the mismatch and match conditions increases (Fig. 1c), consistent with experiments characterizing neural responses during associative learning paradigms [102, 70, 10]. Thus, the presence of the stimulus y suppresses the response evoked by the stimulus x, and can be interpreted as a *prediction* or *expectation* of stimulus x. The amplified response in the mismatch condition is typically interpreted as a *prediction error* signal [114, 239].

We summarized the network responses in a reduced space containing the mismatch responses ($\boldsymbol{r}_x$, $\boldsymbol{r}_y$) and the match response ($\boldsymbol{r}_{xy}$) at different stages of learning (Fig. 1d). We found that $\boldsymbol{r}_x$ and $\boldsymbol{r}_y$ become increasingly *anti*-correlated during learning (Fig. 1e), i.e., the presence of the stimulus y more effectively suppresses the responses to x alone. This anti-correlation does not appear between $\boldsymbol{r}_x$ and $\boldsymbol{r}_y$ of *another stimulus-pair* (Fig. SXXX[1]). Our results suggest that the predictive signal, triggered by the stimulus y, is *specific* to its paired stimulus x, and that the network's predictive computation is not via a global suppression [259]. Although the predictive signal generated by the network is largely driven by the stimulus-specific component, it can also

[1] Show correlation between $\boldsymbol{r}_x$ and $\boldsymbol{r}_y$ for different stimulus-pair doesn't change during learning.

contain a global gain component that is modulated by the total stimulus strengths (SI).

We further found that neural population representations of the match and mismatch conditions *decorrelate* during learning, quantified by the angle between the vectors $\boldsymbol{r}_{xy}$ and $\boldsymbol{r}_x$ (Fig. 1e). This finding is consistent with [70], and suggests that neural responses can be used to effectively decode whether the stimulus x is presented in the match or mismatch condition. Moreover, since each neuron's response is nonlinear with respect to its input, the response in the match condition is not a sum of the responses in the two mismatch conditions (i.e., $\boldsymbol{r}_{xy} \neq \boldsymbol{r}_x + \boldsymbol{r}_y$, Fig. 1c). Importantly, this implies that the network's prediction for the stimuli (x or y) can still be decoded, even though the optimization procedure giving the connectivity of the network did not explicitly require that those readouts are available (Fig. 1a).

Next we make the first step towards the high-dimensional scenario, by examining neural responses to two stimulus-pairs in the match and mismatch conditions ($P = 2$). Ref. [114] proposed that neurons involved in predictive processing are functionally segregated, i.e., neurons that signal prediction error for one stimulus association will tend to signal prediction error for other associations, and similarly for 'representation' neurons that encode the stimulus itself. This proposal has not been tested systematically. When we examined neural responses to two stimulus-pairs in our model, we found no correlations (Fig. 1f, left). This implies, for example, that a neuron that signals prediction error for stimulus-pair 1, may have a selective response to stimulus x 'itself' for pair 2. By contrast, Refs. [114, 184] would predict a high degree of correlation between neural responses in the matched and mismatched conditions to two stimulus-pairs (Fig. 1f). The cellular-level desegregation of response types raises the question of what circuit mechanisms and structure support this functional organization.

4.2.2 Learning and stimulus dimensionality determine the properties of effective predictive processing circuits

We next investigated circuit mechanisms underlying multi-modal high-dimensional predictive processing. To understand why neurons respond more strongly in the mismatch

compared to the match condition, we decomposed the input into each neuron to feedforward and recurrent components, which respectively correspond to the actual stimulus signal, and to internally generated predictions (Fig. 2a, Methods). Such a decomposition is similar to analyses commonly applied to predictive coding experiments, in which the inputs are decomposed to stimulus- and prediction-related components [18, 102, 10].

To quantify the relative contribution of each component, we follow the excitatory /inhibitory (E/I) balance literature [46, 3], and define the *balance* level B as the ratio between the total feedforward input and the net input to each neuron, in each condition (Fig. 2a).

During associative learning, the overall balance level in the network increases in the match condition, and decreases in the mismatch condition (Fig. 2b, left). This is a signature of the increased accuracy with which feedforward sensory input is canceled by the internally generated prediction in the match condition. Notice that the distributions of the balance level (over neurons and stimuli) are initially similar in the match and mismatch conditions, and become significantly different in late stages of learning (Fig. 2b, right). Indeed, in the mismatch condition, the majority of neurons have a balance level close to 0 after training, which explains the network's strong prediction error response (Fig. 1b).

To better understand the role of the balance level in predictive processing, we examined its effect on the nonlinear transformation the network performs, from input stimuli to neural activity (Fig. 2c). Our goal is to identify settings in which the geometry of (potentially noisy) neural responses facilitates robust readout of the prediction error (i.e., whether a particular stimulus was expected or not). We note that this quantity cannot be read-out by a linear classifier from the stimulus input. By contrast, once the input is transformed by into the recurrent network's high-dimensional responses, prediction errors can be linearly read-out. Moreover, while the prediction error itself is stimulus-specific, the decoder that performs this computation is stimulus independent—it is simply the average firing rate (as in Fig. 1b). In other words, the learned structure of neural responses in our model implies that the same linear decoder applies to every stimulus-pair without 're-learning'.

Generally, the distribution of the balance level B depends on multiple external factors, including the stimulus statistics and the frequency with which a stimulus-pair is presented to the animal. We elected to focus on a gain parameter b that controls the slope of each neuron's nonlinear transfer function (Methods, [103]). This parameter is intrinsic to the network, so it could potentially be adjusted dynamically to achieve robust coding. Moreover, there is a direct monotonic relationship between the gain b and the balance level B, which provides means of relating this important model parameter to data.

The parameter b scales both feedforward and recurrent inputs, so increasing it leads to increases of the average firing rates in the match and the mismatch conditions (Fig. 2d, top; Methods). Increasing b also leads to a wider margin between the linear readouts of neural activity in the match and mismatch conditions. Therefore, large b can facilitate decoding the prediction errors, at the cost of increased overall neural activity. These observations suggest together that an intermediate value of b can jointly optimize both encoding efficiency and decoding robustness. As an illustration, we constrain the average network response in the mismatch condition to be larger than a certain threshold, while requiring a minimal but nonzero average response in the match condition (Fig. 2d), consistent with reports of weak neural responses to predictable stimuli [102, 10]. This yields an optimal value, $b^\star$, corresponding to an optimal balance level $B^\star$ (Fig. 2d, bottom). We carried out this optimization procedure for networks that are trained to perform predictive processing of stimulus ensembles of increasing dimensionality (i.e., increasing values of $\alpha = P/N$), with the same constraints on the average firing-rates. This constraint was chosen such that the value of $B^\star$ at $\alpha = 0$ matches experimental data, and on the assumption that an 'over-trained' animal learns a single stimulus-pair (i.e., $\alpha = 1/N \approx 0$). Surprisingly, we found that the optimal balance level *decreases* with α (Fig. 2e). In words, based on our model, we expect that networks that perform predictive processing in natural conditions (large α) should exhibit 'loose' balance. The loose balance, and the corresponding reduced gain, counteracts the interference that arises from learning to generate a large number of internal predictions. We further showed that this prediction of our theory does not depend on other factors such as the

stimulus statistics (Fig. SXXX[2], Methods).

We used neural activity recorded while animals perform learned tasks consisting of visual-motor (V-M) [102] and auditory-motor (A-M) associations [10] to constrain our network model. Specifically, we estimated the balance levels in mouse sensory cortex, by relating the neural data to the model, and assuming that after training the neural network within the mouse cortex reaches the optimal balance level. In the V-M experiment [102], mice were trained to associate their running speed with the speed of visual flow in virtual reality (Fig. 3a). Intracellular voltage levels of primary visual cortex neurons were recorded in the match and mismatch conditions. Fitting the average voltage change in the two conditions to our model and again assuming that $\alpha = 0$ based on the fact that the animals underwent extensive training, the estimated optimal balance level was $B^{\star}_{V-M} = 162 \pm 61$. A consistent result was obtained in the A-M experiment [10], where mice were trained to press a lever and received closed-loop auditory feedback (Fig. 3b,c). Here the recording was extracellular, so balance levels were estimated by equating the ratio of average firing rates in the match and mismatch conditions in the data and model (Methods). Here, we also observed significant variability across animals (Fig. SXXX[3]).

It is notable that the estimate of the balance level is consistent across the two studies with different measurement methods, brain regions, sensory modalities and laboratories. While all these factors may affect the balance level to some degree, our model predicts that the balance level can decrease by up to one order of magnitude when the stimulus dimension increases (Fig. 2e), suggesting that future experiments in naturalistic conditions could be used to confirm our theoretical predictions by revealing loose balance in animals habituated to more rich sensory environments.

[2] Show that how balance level changes as functions of α and μ when adding noise in the stimulus (corresponding to imperfect pairing of x and y in the match condition).

[3] Show that estimated optimal balance levels for different animals.

4.2.3 Stimulus and prediction error representations are desegregated

We now turn to investigating how different functional responses are organized within the network. Previous work has hypothesized that two distinct neural populations exist in predictive processing circuits: (*i*) *internal representation* neurons that 'faithfully' represent external sensory stimuli and encode internal predictions, and *prediction-error* neurons, which signal the difference between the actual stimulus inputs and internal predictions. Given that these neurons selective to those signals exist also in our network model, we wondered whether there it contains functionally segregated neural populations. We adopted the criteria used in experimental work for classifying the internal representation (*R*) and prediction-error (*PE*) neurons ([259, 114], Methods): *R* neurons are those which respond strongly and similarly in match and mismatch conditions, while *PE* neurons are those which respond strongly in the mismatch condition but weakly in the match condition (Fig. 4a). We further distinguish between representation neurons for x and y inputs.

Based on these criteria, we started by studying the case with a single stimulus pair ($P = 1$), and computed the fraction of neurons that are classified as *R* or *PE* neurons at different stages of associative learning (Fig. 4b). As training progresses, the fraction of *PE* neurons increases significantly, consistent with experiments [141, 70], and with the notion that the network learns to 'recognize' the stimulus pairing. The fraction of *R* neurons remains unchanged. We note however that the fraction of *R* neurons exhibits some dependence on the threshold used for categorization (Fig. SXXX[4]) but the increasing fraction of PE neurons is independent of the choice of threshold.

Next we considered the functional types of each neuron, computed separately with respect to two pairs of stimuli. The hypothesis that predictive processing is segregated [114, 184], asserts that if a neuron is a *PE* neuron for stimulus-pair 1, then, if it is active during presentation of stimuli from pair 2, it will likely be categorized as a *PE* neuron with respect to those stimuli too. To test this hypothesis, we computed the joint distribution of neural responses in the four

[4] Show that how fraction of R neurons change when using different thresholds.

relevant conditions (mismatch/match, stimulus-pair 1/stimulus-pair 2), using the optimal balance level fit to intracellular voltage data [102] (Fig. 2e). We then used the aforementioned criteria to classify *R* and *PE* neurons (separately for each stimulus-pair, Methods). We started with the low-dimensional scenario, where the two stimulus-pairs in question are the only stimuli learned by the network (i.e., $P = 2$, $\alpha = P/N \approx 0$). Surprisingly, we found that although many neurons belong to the same functional type with respect to the two stimulus-pairs, approximately 25% of the neurons are in fact *'mixed'*: they are classified as having different functional type with respect to the two stimulus-pairs (Fig. 4c, left).

Furthermore, increasing the dimension of the stimulus the network learns, leads to an increased fraction of mixed neurons (Fig. 4c,d). Intuitively, the looser balance (on average) between feedforward and recurrent inputs in a network with larger stimulus dimension, leads to a broader distribution of balance levels for single neurons (Fig. 2b, Fig. SXXX[5]). That broader distribution, in turn, affords each neuron more flexibility in its response to different stimulus pairs.

It is notable that the fraction of mixed neurons shown here (Fig. 4d) corresponds to two *specific* stimulus-pairs. Indeed, when we considered instead the entire ensemble of stimuli the network learned, we found that most of the neurons are mixed with respect to at least two stimulus-pairs (Fig. SXXX[6]). Thus, contrary to previous hypothesis [114], neurons with mixed representations of stimuli and predictions are common in the network model, especially in the high-dimensional scenario.

We then turned to testing this key prediction of our theoretical model, by looking for signatures of mixed representations of predictions and stimuli in existing experimental data. In our recent work, we recorded primary auditory cortex responses in mice that were trained to make auditory-motor associations [9]. Importantly, unlike typical experiments characterizing mismatch responses, here animals were presented with *probe* auditory stimuli that differ from

[5] Show that too large or too small values of b will generate less mixed neurons (with fixed threshold).

[6] Show that the fraction of neurons that have mixed representation for any two stimulus-pairs.

the *expected* stimulus that was paired with a lever press during training (Fig. 4e). We computed the difference (Δ) between the mismatch and match neural responses (Fig. 3e, bottom), similarly to our analysis of the neural activity in the model (Fig. 4c). Note that for the probe stimuli, 'match' corresponds to a lever press paired with one of the four variants of the learned sound, while 'mismatch' corresponds to responses following a lever press without a sound. We expected Δ values of mixed neurons to lie in the upper left or lower right corners of the plot (similarly to Fig. 4c, blue rectangles). This would correspond to neurons where match and mismatch responses are similar for the expected sound but differ for the probe sound, or vice versa. We quantify the degree of mixing, or desegregation of the predictive representation, by computing the Pearson correlation coefficient of the Δ values corresponding to the expected sound and each probe sound separately (Fig. 4e). We refer to this coefficient as the segregation index, which is close to 1 if the Δ values are strongly correlated between the two stimulus-pairs (expected, probe; see Fig. 1f, right). A segregation index close to 0 means that R and PE responses to different stimulus-pairs are uncorrelated, and the representations of stimuli and predictions are 'maximally mixed'. We additionally computed representational similarity between the expected and probe sounds, as the correlation between neural responses to those stimuli. Crucially, representational similarity was computed on neural responses in a passive condition, not following a lever press [9]. If neurons are segregated into two functional classes, the segregation index should be close to 1 irrespective of the representational similarity. By contrast, we found that the segregation index depends strongly on the representational similarity (Fig. 4f). Specifically, when the expected and probe sounds are similar (Fig. 4e,f, green shades), the segregation index is close to 1, though a random subsampling analysis indicates a statistically significant effect of the representational similarity on the segregation index. When the probe differs from the expected sound more substantially (Fig. 4e,f, orange), the segregation index drops to ~ 0.5. This relation between representation similarity and degree of segregation is consistent with the prediction of our model, with an appropriate level of coding sparsity (Fig. 4f, Methods). The significant dependence of the segregation index on the representational similarity, and the fact that the segregation index is

substantially smaller than 1, suggest that predictive processing is mixed in the mouse auditory cortex. We note that the analysis presented here is an indirect test of the model prediction, that predictive processing is mixed. Indeed, in the model we analyzed responses to two stimulus-pairs that were both learned (Fig. 4c), while in the experiment the animal was only trained on the expected sound. Nevertheless, we believe that the dependence of the smaller segregation index we found for probe stimuli that are dissimilar to the learned sound provides strong evidence against the notion that predictive processing is functionally segregated into separate neural populations. Our model provides a framework that can be used to generate hypothesis that could be tested more directly based on future experiments.

4.2.4 Predictive processing in excitatory–inhibitory networks

Thus far we have focused on relating neural responses in the model to measurements of excitatory neurons' activity [10, 70, 9]. Our results were obtained by analyzing a recurrent network model where each neuron's projections could be both excitatory (E) and inhibitory (I), i.e., the model network does not obey Dale's law. Given the growing literature on the role of inhibitory neurons in computing predictions [88, 87], we sought link our model to experiments more tightly by extending it to a network with separate E and I neurons. We did so by requiring that the activity of E neurons in the E/I network matches exactly the activity of the neurons in the network investigated previously. This guarantees that the E neurons in our E/I network possess the predictive coding properties we studied so far, and opens the door to study the properties and functional role of I neurons. We then 'solved' for those connectivity structures that obey Dale's law (Fig. 5a, Methods). Mathematically, this was done by decomposing the original connectivity matrix to four separate components, representing E$\rightarrow$E, E$\rightarrow$I, I$\rightarrow$E, and I$\rightarrow$I connections. Each connectivity component was guaranteed to have the 'correct' sign by relying on a non-negative matrix factorization procedure (Methods, [73]), combined with previously proposed 'recipes' for modeling E/I networks [81]. The balance level B defined previously based on feedforward and recurrent inputs (Fig. 2), is equal to the stimulus-specific component of the E/I balance in the E/I

networks (Methods).

To summarize, we used two types of constraints to derive an E/I network model that performs high-dimensional predictive processing, namely, matching the activity of the E neurons to the activity of neurons in our previous model, and decomposing the connectivity to preserve the E/I identity of the presynaptic neuron. These constraints did not yield a *unique* connectivity structure. Rather, we found a one-parameter family of connectivity structures that all meet those constraints. This parameter, denoted λ_{EI}, interpolates between two extremes of structured connectivity in the E/I network (Fig. 5b).

In one extreme ($\lambda_{EI} = 0$), inhibition is *'private'*, i.e., each E neuron projects to a single I neuron whose activity is equal to its 'parent' E neuron. This has been an implicit assumption of previous predictive coding models with lateral inhibition [220, 219, 163]. In the opposite extreme ($\lambda_{EI} = 1$), each I neuron receives a large number of excitatory inputs and signals an *'internal prediction'* of one of the stimuli learned by the network.

Varying λ_{EI} from 0 to 1 gives a continuum of inhibitory representations, each with its own properties, which we investigated using the same approach that we applied to predictive representations in E neurons (Fig. 1b-e, Fig. 4b). We started by computing the alignment of the inhibitory response to the stimulus x in the match ($\boldsymbol{r}_{xy}$) and the mismatch condition ($\boldsymbol{r}_{x}$), as a function of the learning stage (Fig. 5c). Before learning ($\mu = 0$), increasing λ_{EI} leads to a marked decrease in the alignment of these inhibitory population responses, quantified by the cosine similarity. After learning ($\mu \approx 1$), increasing λ_{EI} leads to nonmonotonic effect on the alignment. Intriguingly, for $\lambda_{EI} = 1$, after learning the alignment of I responses in the two conditions is larger than that of E responses (Fig. 5c, compare green and black for $\mu = 1$).

We used the differences in I neurons' response properties to estimate the parameter λ_{EI} based on empirical measurements of regular spiking (RS, putative excitatory) and fast spiking (FS, putative inhibitory) neurons. In other words, our modeling framework allowed us to determine the degree to which inhibitory processing in predictive coding circuits is private vs. signals internal predictions. To achieve that, we computed the correlation between auditory cortex

responses in the match and mismatch conditions, separately for RS and FS neurons [10]. We then compared those correlation values obtained before and after learning of an auditory-motor association (Fig. 5d). The pairing between movement and a probe sound (not presented during training) was regarded as *before*-learning and the pairing between movement and the expected sound as *after*-learning (Methods). We found a significant decrease in this correlation during learning for RS neurons, consistent with the corresponding change of correlation in E population responses in the model (Fig. 5c, light blue circles). By contrast, correlation of FS population responses did not change significantly during learning, which rules out small values of λ_{EI}. Moreover, the correlation of population responses after learning was similar for RS and FS neurons, which rules out large values of λ_{EI}. Taken together, our analysis suggests that a value of $\lambda_{EI} \approx 0.6$ best captures the experimental observations.

Given this experimentally-constrained value of λ_{EI}, we used our theory to generate a number of testable predictions for inhibitory representations in predictive coding experiments. First, we have shown that responses of E neurons in the mismatch conditions corresponding to the stimuli x and y (i.e., $\boldsymbol{r}_x$, $\boldsymbol{r}_y$) become significantly anti-aligned during learning (Fig. 1d,e). By contrast, we predict, based on the value $\lambda_{EI} = 0.6$, that anti-alignment of I responses during learning is significantly weaker (Fig. 5e, left). Second, based on the value $\lambda_{EI} = 0.6$, we predict large correlations between inhibitory responses to the stimulus y alone and responses in the match condition (Fig. 5e, middle), when compared with E responses. In the model, the asymmetry of the overlaps $\boldsymbol{r}_x \cdot \boldsymbol{r}_{xy}$ and $\boldsymbol{r}_y \cdot \boldsymbol{r}_{xy}$ arises from the non-negative matrix factorization solution we found when decomposing the connectivity. Although this asymmetry stems from a purely mathematical reason, it may in the future be related to different functional responses of inhibitory neurons of different subtypes [67]. Third, when $\lambda_{EI} = 0.6$, the fraction of I neurons classified as R neurons is less sensitive to the different stages of learning, compared to I neurons in the 'private inhibition' case ($\lambda_{EI} = 0.6$, Fig. 5e, right). We note that the effect of learning on the fraction of E neurons categorized as R neurons depends on the threshold (Fig. SXXX[7]). For this

[7] Show how fraction of R inhibitory neurons when using different thresholds.

reason, it may not be straightforward to relate differences between E and I populations to the network model.

Previous work on predictive coding suggested that associative learning enhances top-down inhibitory projections from outside the local circuit [114, 70]. In this view, after training, inhibitory input that signals the prediction cancels bottom-up excitatory inputs and suppresses neural responses in the match condition. We therefore wondered what changes in inhibitory connectivity during learning lead to stimulus-specific suppression of neural activity in our E/I network model. One option is that inhibitory connections that predict the stimulus are strengthened during learning, as suggested in Ref. [114]. An alternative is that reorganization of inhibition during learning is more subtle, and that inhibitory signals are distributed differently before and after learning. We used our solution to the non-negative matrix factorization problem to calculate the distribution of I-to-E synaptic weights before and after learning. First, we computed this distribution in the scenario of private inhibition ($\lambda_{EI} = 0$, Fig. 5f). The overall distribution is broadened during learning. Examining the change in synaptic weights conditioned on the functional cell-type of pre- and post-synaptic neurons (*R* or *PE*), suggests that stimulus-specific suppression of E responses arises from potentiated I synapses from neurons 'faithfully' representing the stimulus. In other words, when inhibition is private, the predictive signal arises in part due to strengthened projections from inhibitory *R* neurons to excitatory *R* neurons (Fig. SXXX[8]).

Next we examined I-to-E connectivity when inhibitory structure was matched to experimental data ($\lambda_{EI} = 0.6$, Fig. 5g). We found that learning leads to overall sparsification of I connections. Interestingly, here *R*-to-*R* connections can be either potentiated or depressed, unlike the $\lambda_{EI} = 0$ case (compare middle panel of Fig. 5f,g). Moreover, when $\lambda_{EI} = 0.6$, inhibitory connections originating from *PE* neurons that are initially very weak get strongly potentiated.

Together, our results suggest that (*i*) Predictive processing is learned without large changes to the average inhibitory connection strength. This was also seen for other values of

[8] Show that R to PE synaptic weights don't potentiate a lot.

λ_{EI} (Fig. SXXX[9]). (*ii*) The 'strategy' for learning predictive processing can differ substantially, and depend on the underlying circuit structure. This suggests that multiple inhibitory cell-types might be involved in learning of predictive processing based on diverse plasticity rules, and may have different effects on the predictive representations at the network-level. (*iii*) When inhibitory structure is matched to data, the 'internal model' is highly distributed and, surprisingly, arises in part from potentiated connections from inhibitory neurons signaling prediction error. Another signature of this distributed strategy is the fact that in the match condition, the total inhibitory input to each excitatory neuron *decreases* during learning (Fig. SXXX[10]), suggesting that predictive signals are primarily computed by recurrent circuitry rather than directly from top-down inputs.

4.3 Discussion

We presented a study of a recurrent neural network model that learns to generate high-dimensional predictions, generalizing previous theoretical studies, and thus allowing this class of models to be applied in natural conditions. Our analysis of this model gives insight to neural mechanisms supporting high-dimensional predictive coding; generates testable hypotheses for functional properties of biological neural networks; and provides a framework within which experimental data can be quantitatively compared with a normative mathematical model.

We focused on a *recurrent* network model (Fig. 1) for two reasons. First, cortical circuitry that performs predictive processing is known to be highly recurrent. Plasticity of recurrent connections forms functional neuronal assemblies [121], that were suggested to underlie behaviorally-relevant sensory discrimination [19]. Second, predictions for sensory stimuli typically unfold over time, so it is natural to assume that the implementation of these computations is done in a recurrent network based on its intrinsic dynamics [28]. While we focused on steady-

[9] Show that the average inhibitory connection strength doesn't change much during learning for different λ_{EI}'s.

[10] Show that the distribution of inhibitory input magnitudes (sum of all presynaptic inhibitory firing rate $\times$ synaptic weight) becomes smaller.

state neural responses for mathematical tractability, our model could be extended in the future to study the time-dependence of high-dimensional predictive coding. Beyond temporal coding, other interesting theoretical directions to extend our study are: networks with asymmetric connectivity, which could potentially be done by imposing sparsity on the existing connectivity 'prescription' [188]; and networks that learn predictions online, thus different stimuli may be at different stage of learning [240, 188].

Our model suggests that balance between feedforward and recurrent input, or indeed between excitation and inhibition, can lead to robust generation of internal predictions, within the local network. While this has been suggested by previous studies [28, 46, 103], an important novel feature revealed by our analysis is that in realistic conditions there is an optimal, finite balance level, which decreases with stimulus dimension (Fig. 2). Our theory further suggests that the a network with infinitely high balance [46] could be especially vulnerable to noise in high-dimensional scenarios.

Based on our results, we hypothesize that the large degree of heterogeneity of empirical E/I balance levels [3] may be a signature of the differences in the stimulus ensembles animals were exposed to in different experiments. Our results in Fig. 2 and Fig. 3 suggest that this hypothesis could be tested systematically by exposing animals to more rich sensory environments. Here too it may be important to study temporal dynamics of the model, as it was previously shown that synaptic delays (that we ignored) affect the optimal degree of balance in circuits performing low-dimensional predictive coding [103].

The role that balance plays in computing predictions has two important implications for the architecture and function of predictive processing circuits.

- **What is the source of predictive signals?** Previous work has shown that cross-modal predictions are often *stimulus-specific* [259, 70]. More specifically, signals from one brain region can suppress responses to a particular predictable stimuli in another region (e.g., motor cortex activity suppressing visual cortex responses). It is notable that within

our model those computations can be done without fine-tuning the targets of long-range projections, as has been suggested previously [114]. Rather, local recurrent connections in the 'receiving region' can extract the predictions from long-range inputs, relying on E/I balance.

- **What is the timescale of learning predictions?** Prediction error responses in the same cortical region can arise after as little as minutes of training [233] or only after days of training [70, 10]. We believe that the diversity of mechanisms that together enforce E/I balance (e.g., firing-rate adaptation, synaptic-scaling, Hebbian plasticity; see review in Ref. [111]), may explain this wide range of predictive processing learning dynamics. Future work may reveal that our model has explanatory power also for the emergence of predictions over faster timescales than reported in the experiments we considered here.

Given that, these mechanisms enforcing balance identified in broad range of cortical circuits suggests that our theory could be used to understand the predictive processing in subcortical regions and brain circuits in invertebrates.

An important finding based on our model is that predictive representations are *desegregated*: neurons that signal prediction errors for one stimulus-pair may faithfully represent the presence of a stimulus for a second pair. Based on experiments where animals were probed with multiple types of unexpected sounds, we found a signature of this desegregation at the cellular level in mouse auditory cortex (Fig. 4). Our findings are related to the expanding literature on *mixed-selectivity* [198, 68, 156], where neurons exhibit complex tuning to multiple stimulus features. Here we report that neurons may have mixed-selectivity to internally generated predictions of sensory and motor variables (Fig. 4, Fig. 5). There is however one important distinction between our results and the mixed-selectivity literature. In most reported examples of mixed-selectivity, neurons' tuning curves are unstructured [68]. This means that reading out task-relevant variables requires a finely-tuned decoder that is adjusted based on the tuning of the upstream neurons. By contrast, the representations of stimuli in matched and mismatched

conditions in our model is highly structured, and reading out the prediction error can be done without any adjustments to the decoder (Fig. 2). These structured neural response do not arise in other random architectures that give rise to mixed-selectivity.

It is noteworthy that a recent paper has argued that specific genetic markers are over-expressed in neurons that belong to specific functional classes (in the predictive coding framework, [184]). The results in that study therefore appear to be in tension with our model and its interpretation presented here. We think however that there is no contradiction, based for the following reasons:

- According to the method in Ref. [184], neurons are photo-converted ('captured') based on expression of a single genetic marker, while the relevant markers can be co-expressed [229]. If the interpretation offered in [184]—that genetic expression determines function—is correct, then co-expression of multiple markers and varying levels of each individual marker may in fact be a signature of desegregated predictive representations. Indeed, the link Ref. [184] establishes between the genetic markers and functional cell-types relies on binary classification to neurons with 'high' expression. In our view, functional properties of neurons are expected to lie on a continuum (Fig. SXXX[11]) which may be related to the continuous expression levels of the relevant genetic markers.

- In Ref. [184], classification of captured neurons into cell-types requires alignment with the Allen Institute Atlas [229]. We believe that this alignment step may introduce artifacts that make it difficult to definitively identify classify the captured neurons. Additionally, the small number of animals used in [184] may imply that across neuron variability was overestimated, since expression profiles of different neurons in the same animal could be correlated.

In summary, predictive processing is a ubiquitous and fundamental computation support-

[11] Show that the distribution of number of stimuli for each neurons such that it is a R or PE or either neuron; and joint distribution of them for the number of R and PE stimulus-pairs.

ing diverse behaviors across animal species. Experimental measurement of internally generated predictive signals is especially challenging, because it requires identifying the source of inputs to each neuron [239]. We provided a mathematical framework that can be used to interpret neural recordings during behavior in naturalistic sensory environments and decipher the circuit mechanisms underlying predictive processing. Beyond comparison with experimental data, we believe that our work will advance the understanding of how the brain constructs complex internal-models by shedding light on commonalities and differences between biological predictive coding circuits and artificial systems, particularly those trained using self-supervised algorithms [140].

4.4 Methods

Recurrent network model

Our model network consists of N neurons whose firing rates are described by the time-dependent vector $\boldsymbol{r}(t) = (r_1(t), \ldots, r_N(t))$. The network is driven by high-dimensional stimulus input, denoted $\boldsymbol{x}(t) = (x^1(t), \ldots, x^P(t))$ and $\boldsymbol{y}(t) = (y^1(t), \ldots, y^P(t))$. The vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ correspond to stimuli from two modalities that are paired during training.

The dynamics of the recurrent network are given by

$$\frac{dh_i(t)}{dt} = -h_i(t) + \underbrace{b\sum_{j=1}^{N} J_{ij}\phi(h_j(t))}_{-I_i^R} + b\,I_i^F(\boldsymbol{x}(t), \boldsymbol{y}(t)). \tag{4.1}$$

Here $h_i(t)$ is the voltage level of each neuron and is related to its firing rate via a nonlinear activation function, $r_i(t) = \phi(h_i(t))$. Note that the input each neuron receives is decomposed in Eq. (4.1) into the recurrent (I_i^R) and feedforward (I_i^F) components. We rescaled the connectivity matrix J_{ij} and the feedforward input $I_i^F(\boldsymbol{x}(t), \boldsymbol{y}(t))$ by a constant b, which can be interpreted as a gain parameter. The explicit forms of J_{ij} and $I_i^F(\boldsymbol{x}(t), \boldsymbol{y}(t))$ were determined based on a normative approach. We assume that the neurons' dynamics jointly minimize the following

objective

$$E(t) = \underbrace{\sum_{k=1}^{P}\left[\left(x^k(t+d)-\hat{x}^k(t)\right)^2+\left(y^k(t+d)-\hat{y}^k(t)\right)^2\right]}_{\text{Prediction errors}} + \underbrace{\frac{2}{b}\sum_{i=1}^{N}F(r_i(t))}_{\text{Encoding efficiency}}, \tag{4.2}$$

where $\hat{x}(t)$ and $\hat{y}(t)$ are the internal predictions generated by the network at time t and $F(r)$ is a monotonically increasing function whose explicit form depends on ϕ, the nonlinear activation function (SI). For ReLU nonlinearity [$\phi(z) = \max(z-\theta, 0)$], the appropriate choice is $F(r) = \frac{(r+\theta)^2}{2}$.

We further assume that the internal predictions are linear readouts of the network activity

$$\hat{x}^k(t) = \frac{1}{N}\boldsymbol{w}^k \cdot \boldsymbol{r}(t), \qquad \hat{y}^k(t) = \frac{1}{N}\boldsymbol{v}^k \cdot \boldsymbol{r}(t). \tag{4.3}$$

Here $\boldsymbol{w}^k, \boldsymbol{v}^k \in \mathbb{R}^N$ are the readout weight vectors. These internal predictions are, by definition, predictions of future input, as indicated by the delay d in Eq. (4.2). However, we will focus on the scenario where the input changes much more slowly than the neurons' firing rates. Therefore, on the timescale of firing rate changes, we will regard the stimulus inputs to be approximately constant, i.e.,

$$x^k(t+d) \approx x^k(t) \approx x^k, \qquad y^k(t+d) \approx y^k(t) \approx y^k. \tag{4.4}$$

We note that the parameter b in Eq. (4.2) controls a trade-off between minimizing the prediction error and maximizing the encoding efficiency.

For mathematical tractability, the weight vectors $\boldsymbol{w}^k$ and $\boldsymbol{v}^k$ [Eq. (4.1)] are assumed to be correlated random variables. Specifically, each component of $\boldsymbol{w}^k$ and $\boldsymbol{v}^k$ is a Gaussian random

number with mean zero and unit variance, and the correlation between them is μ^k,

$$\begin{aligned} \langle w_i^k \rangle = \langle v_i^k \rangle &= 0, \\ \langle (w_i^k)^2 \rangle = \langle (v_i^k)^2 \rangle &= 1, \\ \langle w_i^k v_i^k \rangle &= \mu^k. \end{aligned} \tag{4.5}$$

The correlation of each stimulus-pair, μ^k, increases during learning. For simplicity, unless noted otherwise, all stimulus-pairs have the same 'age', i.e., μ^k does not depend on the index k. Under these assumptions, we derived the dependence of the steady-state firing rate distribution on the stimuli presented to the network and the correlation μ in the limit where both N and P are large, and their ratio $\alpha = P/N$ is nonzero (SI).

The presence or absence of each stimulus was modeled by setting the corresponding components of $\boldsymbol{x}$ and $\boldsymbol{y}$ to 0 or 1. For example, the mismatch and conditions for k-th stimulus-pair correspond to,

$$\begin{aligned} &(x^k, y^k) = (1,0) \qquad (x\text{-only mismatch condition}), \\ &(x^k, y^k) = (0,1) \qquad (y\text{-only mismatch condition}), \\ &(x^k, y^k) = (1,1) \qquad (\text{match condition}) \end{aligned}$$

Definition of balance level

The balance level for neuron i is defined as,

$$B_i = \left| \frac{I_i^F}{I_i^F - I_i^R} \right|. \tag{4.6}$$

Here, I_i^F and I_i^R are the feedforward and recurrent input currents to neuron i at steady-state [Eq. (4.1)]. The balance level varies between neurons and between stimuli, because the weights w_i^k and v_i^k are different for different neurons and stimuli (indexed by i and k, respectively). The

balance level distribution and its median shown in Fig. 2 were computed analytically (SI).

Extracting the optimal balance level from data

V-M experiment, Ref. [102]. We calculated the trial-averaged voltage of all the recorded L2/3 neurons as a function of time (Fig. 2a). Voltage level of each neuron was measured with respect to its baseline. We sampled 50 voltage levels from all recorded neurons and all time points in the match and mismatch time windows (Fig. 2a), which were $-0.1-0$s (match) and $0-0.1$s (mismatch). The time $t=0$ corresponds to point at which the treadmill was decoupled from visual flow in virtual reality. We then computed the standard deviation over those 50 samples of the voltage level in the match and mismatch conditions. By taking the ratio of these standard deviations, we obtained a dimensionless quantity that has a direct analogue in the model. We fit the model to this ratio by adjusting the value of b. The best-fit value $b^\star$ gives the median of balance level $B^\star$ in the network model.

A-M experiment, Ref. [10]. We calculated the trial-averaged firing rates for all regular spiking neurons ($n=815$) in the movement and active condition in two time windows: from $t=-0.1$s to stimulus onset ($t=0$), and from stimulus onset to $t=0.06$s (Fig. 3b). For every neuron, we calculated the change in its firing rate between the two time windows in both conditions. We sampled 400 firing rate change values from 815 neurons with replacement, and calculated the average firing rate change in the passive (mismatch) and movement (match) conditions. The model was fit by finding the parameter b for which the ratio of the average firing rate change in the two conditions was equal in the data. Again we calculated the median of balance level $B^\star$ based on the best-fit value of b.

The fitting procedure for both experiments was repeated 100 times, giving the estimated values of $B^\star$ shown in Fig. 3c.

Definition of functional cell types

We denote the steady-state voltage of neuron i in the mismatch conditions as h_i^x (x-only) and h_i^y (y-only), and in the match condition as h_i^{xy}. To classify neurons into functional types, deviations of individual neurons' voltage response relative to the mean were compared to the standard deviation (denoted σ) of the steady-state voltage distribution. We evaluated σ using the voltage distribution in the x-only mismatch condition after learning ($\mu = 0.97$). We note that in the network model the average voltage response for any μ and in any condition is 0, which simplifies the criteria below.

A neuron i is a representation (R) neuron for the x-stimulus if it is depolarized upon presentation of the stimulus x, i.e., its voltage response in x-only mismatch condition is large, and its voltage responses in the match and mismatch conditions are similar. Mathematically,

$$h_i^x > \frac{\sigma}{2} \quad \text{and} \quad \left|h_i^x - h_i^{xy}\right| < \frac{\sigma}{2}. \tag{4.7}$$

A similar criterion was used to identify R neurons for the y-stimulus. A neuron i is a prediction error (PE) neuron if it signals the 'mismatch' between x and y, i.e., its voltage response in the x-only mismatch condition is large, and its voltage response in the match condition is small. Mathematically,

$$h_i^x > \frac{\sigma}{2} \quad \text{and} \quad h_i^x - h_i^{xy} > \frac{\sigma}{2}. \tag{4.8}$$

Neurons meeting these criteria are referred to as *positive* PE neurons, because their activity increases when x is presented but not expected (based on y). The activity of *negative* PE neurons increases when x is not presented but is expected.

Note that neurons in the network may not belong to any of the those three classes. The fraction of R and PE neurons in the network for different values of μ and α, shown in Fig. 4b,d, were computed analytically (SI). Other possible functional classes of neurons are discussed in the SI.

Estimating functional segregation from responses to multiple stimuli

We calculated the trial-averaged firing rate change of each neuron in the match (movement) and mismatch (passive) conditions, separately for each sound stimulus [9]. For each type of probe sound, we restrict the analysis to neurons responsive in the passive condition to that probe sound and the learned (expected) sound. The firing rate threshold we used to identify responsive neurons was half standard deviation above the mean firing rate for expected sound in the passive condition. Changing the threshold does not affect the results in Fig. 4e,f.

For these neurons, we computed pairs of Δ values, defined as the difference between mismatch and match responses, for the probe and expected stimulus. The Pearson correlation coefficient between those Δ values was defined as the segregation index.

To estimate the similarity of the expected and probe stimuli, we computed individual neurons' trial-averaged firing rate change following presentation of those stimuli in the passive condition (the same time windows used in the A-M experiment as in Fig. 3). For each animal, we considered population vectors consisting of all its recorded neurons. Representation similarity was defined as the Pearson correlation of those vectors for pairs of auditory stimuli (expected and probe, Fig. 4f). We note that to calculate this similarity in the model is calculated from the activity of all neurons that are active in either the expected or probe stimuli in passive condition.

E/I network model

In the network with separate E and I, the time dependent voltages of E and I neurons are given by the following set of differential equations,

$$\begin{aligned}
\frac{dh_i^E}{dt} &= -h_i^E + \sum_{j=1}^{N_E} J_{ij}^{EE}\phi(h_j^E) + \sum_{j=1}^{N_I} J_{ij}^{EI}\phi_I(h_j^I) + I_i^E, \\
\tau_I \frac{dh_i^I}{dt} &= -h_i^I + \sum_{j=1}^{N_E} J_{ij}^{IE}\phi(h_j^E) + \sum_{j=1}^{N_I} J_{ij}^{II}\phi_I(h_j^I) + I_i^I.
\end{aligned} \tag{4.9}$$

We assume that the activation function for inhibitory neurons is ReLU with zero threshold, $\phi_I(x) = \max\{x, 0\}$. Matching the E neurons' activity at steady state to the activity of neurons in our original network [Eq. (4.1)] gives constraints on the connectivity components and the feedforward input (SI),

$$\begin{aligned} J^{EE} - J^{EI}(I + J^{II})^{-1} J^{IE} &= J, \\ \boldsymbol{I}^E - J^{EI} \boldsymbol{I}^I &= \boldsymbol{I}^F. \end{aligned} \tag{4.10}$$

Here J and $\boldsymbol{I}^F$ are the connectivity matrix and feedforward input used in Eq. (4.1). We further assume that the matrix $I + J^{II}$ is invertible. In general, there are many possible solutions $\{J^{EE}, J^{EI}, J^{IE}, J^{II}, \boldsymbol{I}^E, \boldsymbol{I}^I\}$ satisfying Eq. (4.10). We identify a family of solutions. This continuum interpolates between the solution with private inhibition, where J^{IE} is equal to the identity matrix; and solutions with an inhibitory internal prediction, where rows of J^{IE} are given by the stimulus weight vectors (SI). Moreover, we show that up to a constant, this balance level is the same as the stimulus-specific, local component of the E/I balance level in the E/I network (SI).

We extended the definition of functional cell-types [Eqs. (4.7,4.8)] to I neurons. We note that here the average input to inhibitory neurons is not 0, so we subtracted the mean from the voltage level [h's in Eqs. (4.7,4.8)] before applying the criteria on the deviations from the mean.

Analyzing responses of regular spiking and fast spiking neurons

We estimated the connectivity structure parameter λ_{EI} based on recordings of regular spiking and fast spiking neurons [10]. Using the same time windows as Fig. 3b and Fig. 4e,f, we calculated individual neurons' trial-averaged firing rate change in the passive and movement conditions for the expected sound and a probe sound. Those firing rate changes recorded in each animal form eight population vectors (regular/fast spiking, expected/probe sound, active/passive). We computed the Pearson correlation between the population vectors in the active and passive conditions, giving four values for each animal, shown in Fig. 5d. The correlation values for

presentation of the expected sound were regarded as ‘after learning’, while correlation values for presentation of a probe sound that was not associated with the lever press were regarded as ‘before learning’.

Statistical tests

In Figs. 3c, 4f and 5d, we used two-sided, unpaired t-tests. $^{\star} = p < 0.05$ and $^{\star\star\star} = p < 0.0005$.

Acknowledgment

Chapter 4, in part is currently being prepared for submission for publication of the material. Wang, Audette, Schneider and Aljadeff. ”Desegregation of neural predictive processing”. The dissertation author was the primary author of this material.

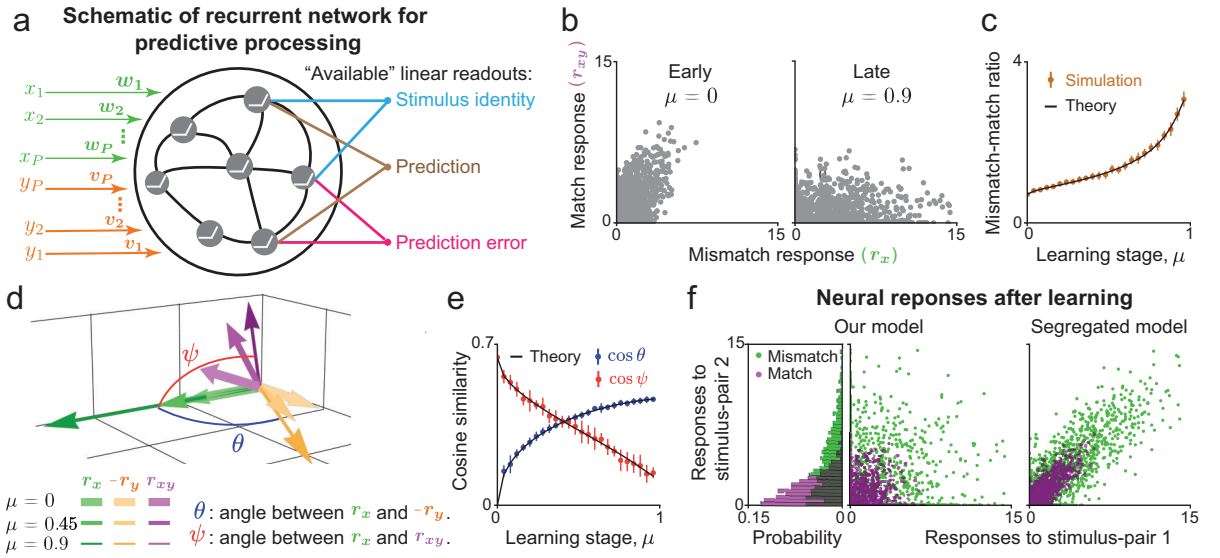


Figure 4.1. Emergence of predictive stimulus representations in a recurrent network model during learning. (a) Schematic of a recurrent network model driven by P pairs of stimuli (x and y). Associative training increases the correlations between the feedforward weights carrying the input signals ($\boldsymbol{w}$ and $\boldsymbol{v}$). The recurrent weights jointly minimize prediction errors and overall encoding efficiency. The neural representation formed under such optimal recurrent connectivity allows flexible linear readouts of various stimulus features. (b) Firing rate responses of individual neurons in the match and mismatch conditions. Initially match and mismatch responses are correlated. After learning, responses are not correlated, and match responses are suppressed. (c) The ratio between average firing rates in the mismatch and match conditions increases during learning. (d) Reduced three-dimensional neural activity space. Each vector represents the mean-subtracted firing rate vector of neurons in the network at different conditions and stages of learning. (e) Learning leads to anti-correlation between neural responses to the stimuli x and y when presented separately, and decorrelates the neural responses in the match and mismatch conditions. (f) Firing rate responses of individual neurons to two stimulus-pairs in the match and mismatch conditions. In our model (left) there are no correlations between the responses to the two stimuli. Those responses are expected to be strongly correlated in a model in which predictive coding is functionally segregated (right). Error bars indicate standard deviations over 10 samples of the network.

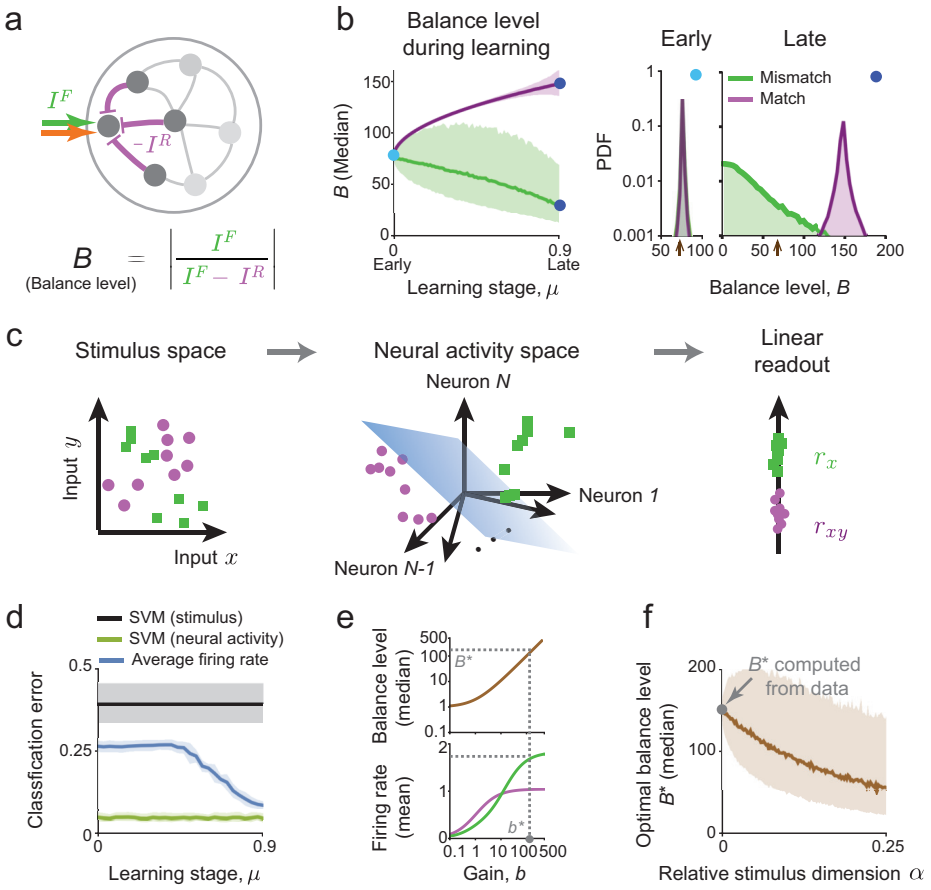


Figure 4.2. Balance between feedforward and recurrent inputs is an important mechanism supporting predictive processing. (a) The input to each neuron is decomposed into feedforward and recurrent components. Each neuron's balance level B is the ratio the total feedforward input and the net input. (b) The median of B in the match and mismatch conditions during learning (left, shaded area indicates inter-quartile range). 'Snapshots' of the distributions of B early and late in learning show how the match and mismatch conditions become separable (right). (c) Schematic showing the nonlinear transformation from the stimulus space (left) to neural activity space (center), which facilities a linear read-out of relevant stimulus features (here, presentation of x in match/mismatch condition). (d) Error of a support vector machine classifier trained to identify the condition based on the input stimuli (black) and on neural responses (green). After learning, a linear classifier based on the average firing rate (blue) performs almost as well as the optimal classifier, suggesting that functionally relevant features from all stimulus-pairs can be extracted without re-learning. (e) The balance level B increases monotonically with the gain parameter b (top). Increasing b leads to a larger margin between match and mismatch responses (improved separability) at the cost of higher firing rates (bottom). The optimal balance level $B^\star$ is determined by constraining the average firing rate in the mismatch condition and minimizing it in the match condition. (e) Increasing the stimulus dimension leads to decrease in $B^\star$, i.e., a more loose balance. At $\alpha = 0$, we fit $B^\star$ to experimental data (Methods).

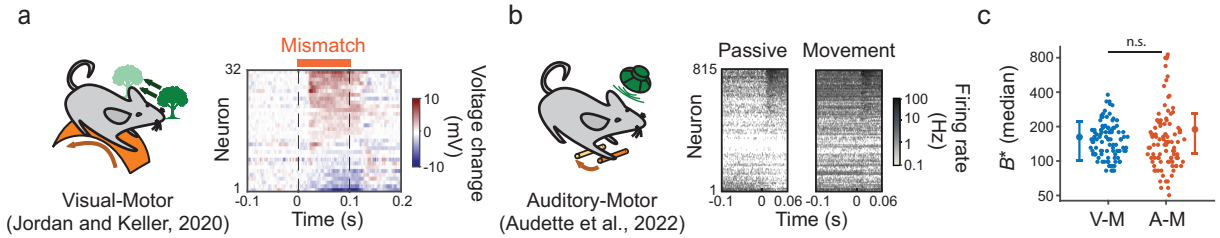


Figure 4.3. Estimating the balance level from predictive coding experiments. (a) Schematic of a learned visual-motor association between running and virtual reality visual flow [102]. Intracellular voltage levels of different neurons in primary visual cortex reveal tuning to mismatch between running speed and visual flow (prediction errors). (b) Schematic of a learned audio-motor association between a lever press and a sound [10]. Extracellularly recorded firing rates reveal tuning to sound omission and to auditory stimuli presented without a lever press (prediction errors). (c) Estimating the median optimal balance level for V-M (blue) and A-M (orange) experiments gives similar values. Error bars are based on repeated subsampling (Methods).

Figure 4.4. (Next page) Desegregated stimulus and error representations in networks performing high-dimensional predictive processing. (a) Schematic of typical tuning profiles of different functional cell-types to the stimuli x and y. (b) Fraction of representation (R) and prediction error (PE) neurons in the model at different learning stages. Error bars indicate standard deviation over 10 instances of the network. (c) Joint distribution of individual neurons' Δ values , the difference between the neuron's , mismatch and match responses, for two stimulus-pairs. Only neurons responsive to both stimulus pairs are included in the distribution (Method). Mixed representation neurons have significantly different Δ values for the two stimulus-pairs, i.e., they are in the blue rectangular regions. As the stimulus dimension (α) increases, more neurons have a mixed representation of the stimulus and prediction error. (d) The fraction of mixed representation neurons increases as stimulus dimension increases. Error bars indicate standard deviations for 200 instances of the network. (e) Evaluating the segregation of stimulus and prediction error representations based on neural recordings during a learned auditory-motor association. Shown are the Δ values of stimulus-responsive neurons for the expected sound and each probe type (colors). Red ellipses indicate the spread of data. The length and direction major/minor axis shows the amplitude and direction of the two principle components. (f) Segregation index as a function of representation similarity for different pairs of expected and probe sounds. Colored points correspond to subsamples of the data, and crosses correspond to the average for each probe type (Methods). Experimental data is compared with equivalent quantities from the model, obtained by varying the sparsity of responses in the model (f, see SI for details).

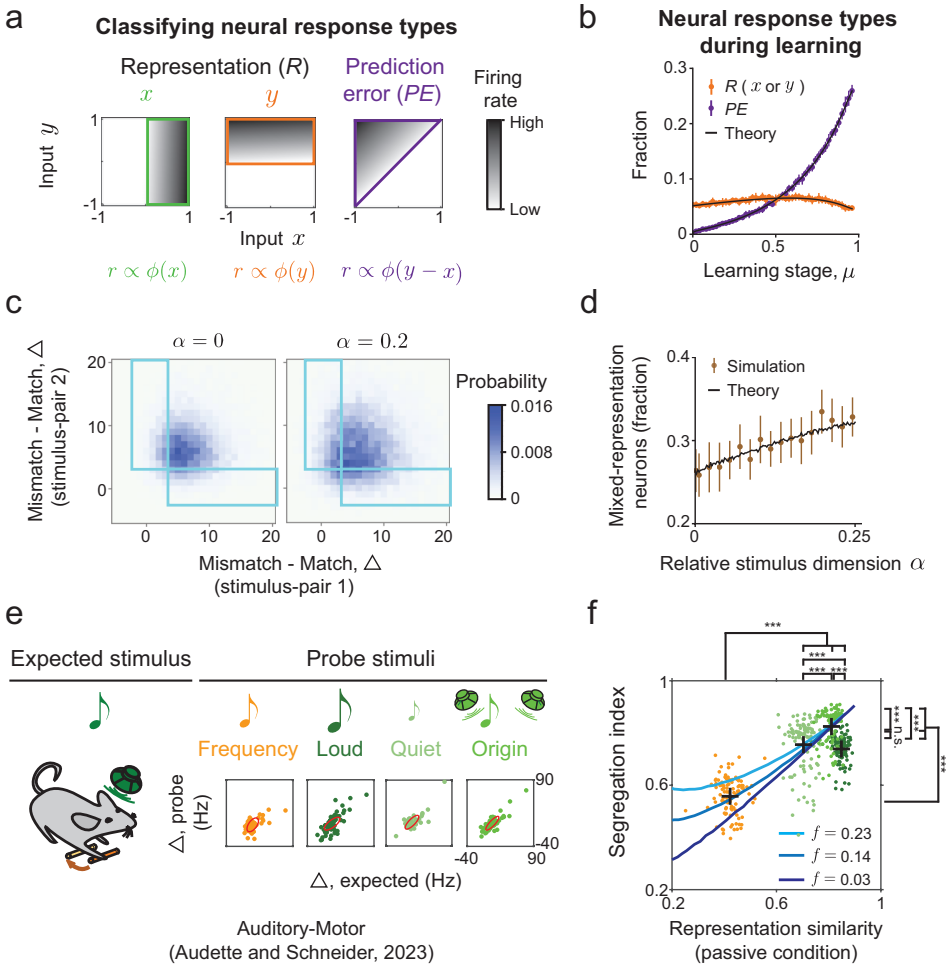
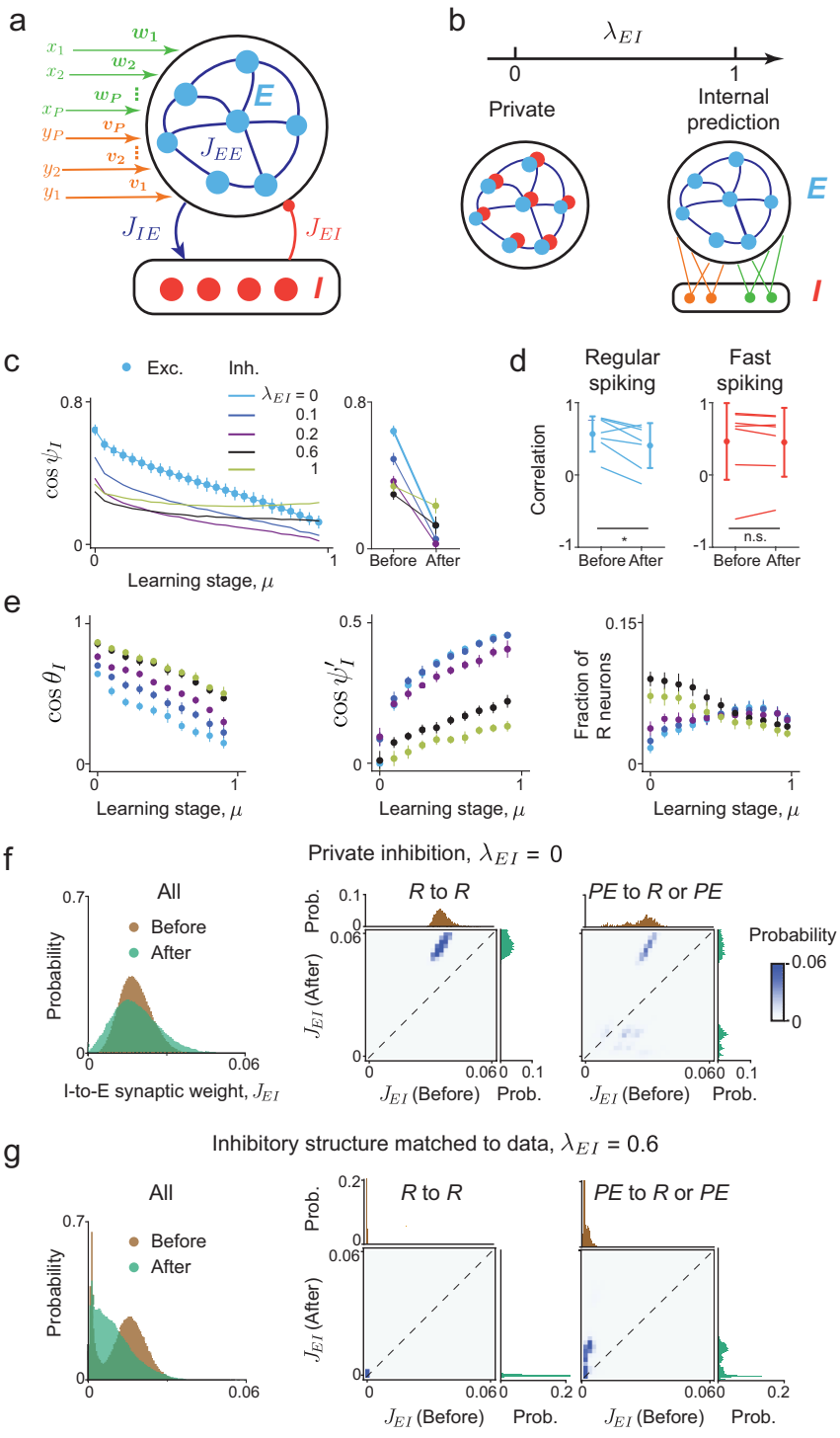

a
Classifying neural response types
Representation (R)
Prediction error (PE)
Firing rate
High
Low
Input y
Input x
r ∝ φ(x)
r ∝ φ(y)
r ∝ φ(y − x)
b
Neural response types during learning
R (x or y)
PE
Theory
Fraction
Learning stage, μ
c
α = 0
α = 0.2
Probability
Mismatch - Match, △ (stimulus-pair 2)
Mismatch - Match, △ (stimulus-pair 1)
d
Simulation
Theory
Mixed-representation neurons (fraction)
Relative stimulus dimension α
e
Expected stimulus
Probe stimuli
Frequency
Loud
Quiet
Origin
△, probe (Hz)
△, expected (Hz)
Auditory-Motor
(Audette and Schneider, 2023)
f
Segregation index
f = 0.23
f = 0.14
f = 0.03
n.s.
Representation similarity
(passive condition)

Figure 4.5. (Next page) A data-constrained excitatory/inhibitory model suggests that internally-generated predictions are distributed across the network. (a) Schematic of the E/I network with separate connectivity components. Excitatory neurons receive external inputs, and their activity is constrained to equal that of neurons in our original model. (b) A family of E/I networks that satisfy the desired constraints, identified based on non-negative matrix factorization. Solutions are parameterized by λ_{EI}, which interpolates between 'private' inhibition and inhibition that signals 'internal predictions'. Varying λ_{EI} gives different patterns of inhibitory responses and connectivity structures. (c) The angle ψ_I between the match and mismatch inhibitory population responses to stimulus x ($\boldsymbol{r}_{xy}$, $\boldsymbol{r}_x$), for different values of μ and λ_{EI} (left). Comparing $\cos\psi_I$ before and after learning (right) allowed us to link inhibitory connectivity structure to inhibitory representations. (d) Analogous correlation between empirical population responses, computed separately for regular- and fast-spiking neurons. Each point represents data from one animal. Correlation mean and standard deviation across animals are also shown. Regular-spiking neurons significantly decorrelate during learning, while fast-spiking neurons' correlation does not change. Correlations of regular- and fast-spiking neurons after learning are similar. (e) The angle θ_I (left) between inhibitory population responses to the paired stimuli in the mismatch conditions ($\boldsymbol{r}_x$, $-\boldsymbol{r}_y$), and the angle ψ_I' (center) between match and mismatch inhibitory population responses to stimulus y ($\boldsymbol{r}_{xy}$, $\boldsymbol{r}_y$). Angles are shown as a function of μ and λ_{EI}, leading to experimentally testable predictions pertaining to inhibitory representations. (right) Fraction of inhibitory R neurons as a function of μ and λ_{EI}. For the experimentally constrained parameter λ_{EI} this fraction decreases for inhibitory neurons (black), while it does not change significantly for excitatory neurons (blue, Fig. 4b). (f) Synaptic weight distribution of all I-to-E connections before and after learning, when $\lambda_{EI} = 0$ (left), and for pairs of E and I neurons belonging to specific functional classes (R to R, middle; PE to R or PE, right). Learning broadens the overall synaptic weight distribution, potentiates the inhibitory connections between inhibitory R neurons. (g) Same as (f), when inhibitory structure is matched to data ($\lambda_{EI} = 0.6$). Here learning sparsifies and depresses inhibitory connections. Connections between R neurons remain very small throughout learning. Surprisingly, connections from inhibitory PE neurons are strongly potentiated.


a
b
c
d
e
f
g
E
I
Private
Internal prediction
Exc.
Inh.
Learning stage, μ
Regular spiking
Fast spiking
Correlation
Before
After
n.s.
Fraction of R neurons
Private inhibition, λEI = 0
All
R to R
PE to R or PE
Probability
I-to-E synaptic weight, JEI
JEI (Before)
JEI (After)
Prob.
Inhibitory structure matched to data, λEI = 0.6

Chapter 5

Discussion and Perspective

Learning is a multifaceted phenomenon. Many aspects of the works in the preceding three chapters can be extended. Here I will discuss three major directions and challenges when extending them to understand other aspects of learning in the brain.

5.1 Learning in neural-network-like architectures

In chapter 4, we formulated a class of recurrent networks that performs high-dimensional predictive processing. But we do not know whether there are also other possible networks with different connectivity that perform predictive processing equally well and whether they exhibit similar mixed representation as in ours. This is an issue both in the original network (with no separate E/I neurons) and in the E/I network. Understanding the general mapping from network structure to its function has been challenging [138, 182], as the space of all possible networks can be huge and intractable to analysis. Recently, many works have been able to identify network solutions that perform specific tasks by optimizing the network connectivity to maximize the task performance [250, 14, 251]. It is found that these trained networks exhibit similar properties and neural representations as observed in experiments. Similar methods might be also helpful to find other possible network solutions to predictive processing.

One issue is that most of these works use stochastic gradient descent (SGD) to find the network connectivity, which is unlikely the algorithm implemented by the brain. On the

other hand, experiments have discovered different types of plasticity rules in different regions in the brain [139]. These plasticity rules can serve as learning mechanisms to tune the network connectivity in the brain. This motivates the general problem on understanding the effects of learning in neural network type architectures. These results are expected to help identify circuit mechanisms on many cognitive tasks.

Learning in these neural network models have been extensively studied in statistical learning community [17, 22, 49]. Yet they mostly focus on how stochastic gradient descent (SGD) enables learning in feedforward network. These networks have been suggested operating at an overparametrized regime where the landscape has many flat minima. The learning algorithm "bias to" those minima that generalizes well in real data [218]. Even if these results are exclusively for artificial networks trained with SGD, the mathematical tools and concepts are expected to be also applicable to rate networks under other learning rules. It would be interesting to see how these types of theory can be built for biological learning rules in the neural networks, such as what is the relevant tangent kernel and implicit bias for biological plasticity rules and how they are different from SGD.

A further extension to these works would be to understand learning and plasticity dynamics in spiking neural networks. This will help a more systematic study of the effect of calcium-based STDP rule as presented in chapter 3. While the available mathematical tools for spiking networks are not as extensive as those for rate-based models, recent studies have begun to pave the way, developing the necessary methodologies [178, 34].

5.2 Learning across multiple timescales

In chapter 3, we studied spiking network dynamics under calcium-based long-term plasticity. The long-term plasticity ie also the primary mechanism to generate experience-dependent predictions as those studied in chapter 4. While (long-term) synaptic plasticity can span over minutes to days, it only occupies a small portion of the temporal spectra on which

learning mechanisms in the brain can operate. Other learning mechanisms include gating and adaptation on faster timescales, meta-plasticity and evolution on slower timescales etc. For example, various neuromodulators can change the plasticity curves induced by calcium and serve as gating or meta-plasticity mechanisms [157, 36]. Short-term plasticity can facilitate stimulus-specific adaptation [86, 233] for predictive processing at much faster timescales.

It is believed that such diversity and heterogeneity in learning mechanisms is crucial to the brain's ability of lifelong learning [130]. How do different mechanisms collectively support learning while not interfering with each other? Similar question arises in machine learning community, known as continual learning. Addressing this question may require a better understanding of how different task information are organized within the same network during learning, which would also advance the challenges mentioned in the last section.

5.3 Learning in higher-level cognition: abstraction, mathematical reasoning and more.

In chapter 4, we studied the high-dimensional predictive representation in recurrent networks. Learning helps the network form a better and better reconstruction of the multi-sensory inputs under the energy constrain. Interestingly, during the same learning process, the network also learns to represent an abstract, categorical variable, which encodes whether the sensory inputs are familiar or not (Figure 4.2cd). This is a simple example of abstraction that can be learned by the circuits in the brain.

Decoding the neural mechanisms behind higher-order cognitive capability, like abstraction and reasoning, is arguably one of the greatest challenge of neuroscience. Abstraction, together with mathematical reasoning, forms the basis of human capacity for performing scientific research. Although the mental mechanisms underlying these high-order cognition have been concerned by the psychology and cognitive science community for decades [137, 30], no current artificial systems or network models can autonomously accomplish these tasks. This is

likely because we lack a good understanding of the related functioning principles. Unlocking these principles would certainly be vital for both neuroscience [136] and artificial intelligence [44, 242]. Even if building such a system in reality necessitates future technological enhancements in hardware or computational power, these principles will help to elucidate the cognitive limits of the brain.

Bibliography

[1] The Supplemental Material is available in the online version of the paper.

[2] Claudio Acuna, Qingchen Guo, Jacqueline Burré, Manu Sharma, Jianyuan Sun, and Thomas C Südhof. Microsecond dissection of neurotransmitter release: Snare-complex assembly dictates speed and ca2+ sensitivity. *Neuron*, 82(5):1088–1100, 2014.

[3] Yashar Ahmadian and Kenneth D Miller. What is the dynamical regime of cerebral cortex? *Neuron*, 109(21):3373–3391, 2021.

[4] Alan Eric Akil, Robert Rosenbaum, and Krešimir Josić. Balanced networks under spike-time dependent plasticity. *PLoS Computational Biology*, 17(5):e1008958, 2021.

[5] Abdul Rasheed A Alabi and Richard W Tsien. Synaptic vesicle pools and dynamics. *Cold Spring Harbor perspectives in biology*, 4(8):a013680, 2012.

[6] Johnatan Aljadeff, Maxwell Gillett, Ulises Pereira Obilinovic, and Nicolas Brunel. From synapse to network: models of information storage and retrieval in neural circuits. *Current Opinion in Neurobiology*, 70:24–33, 2021.

[7] Johnatan Aljadeff, Merav Stern, and Tatyana Sharpee. Transition to chaos in random networks with cell-type-specific connectivity. *Physical Review Letters*, 114(8):088101, 2015.

[8] Christina Allen and Charles F Stevens. An evaluation of causes for unreliability of synaptic transmission. *Proceedings of the National Academy of Sciences*, 91(22):10380–10383, 1994.

[9] Nicholas J. Audette and David M. Schneider. Stimulus-specific prediction error neurons in mouse auditory cortex. *Journal of Neuroscience*, 2023.

[10] Nicholas J Audette, WenXi Zhou, Alessandro La Chioma, and David M Schneider. Precise movement-based predictions in the mouse auditory cortex. *Current Biology*, 32(22):4925–4940, 2022.

[11] Norbert Babai, Olexiy Kochubey, Daniel Keller, and Ralf Schneggenburger. An alien divalent ion reveals a major role for ca2+ buffering in controlling slow transmitter release. *Journal of Neuroscience*, 34(38):12622–12635, 2014.

[12] Craig H Bailey, Eric R Kandel, and Kristen M Harris. Structural components of synaptic plasticity and memory consolidation. *Cold Spring Harbor perspectives in biology*, 7(7):a021758, 2015.

[13] Richard W Baker and Frederick M Hughson. Chaperoning snare assembly and disassembly. *Nature reviews Molecular cell biology*, 17(8):465, 2016.

[14] Omri Barak. Recurrent neural networks as versatile tools of neuroscience research. *Current opinion in neurobiology*, 46:1–6, 2017.

[15] Jérémie Barral and Alex D Reyes. Synaptic scaling rule preserves excitatory–inhibitory balance and salient neuronal network dynamics. *Nature Neuroscience*, 19(12):1690–1696, 2016.

[16] Ellen F Barrett and CF Stevens. The kinetics of transmitter release at the frog neuromuscular junction. *The Journal of physiology*, 227(3):691–708, 1972.

[17] Peter L Bartlett, Andrea Montanari, and Alexander Rakhlin. Deep learning: a statistical viewpoint. *Acta numerica*, 30:87–201, 2021.

[18] Andre M Bastos, W Martin Usrey, Rick A Adams, George R Mangun, Pascal Fries, and Karl J Friston. Canonical microcircuits for predictive coding. *Neuron*, 76(4):695–711, 2012.

[19] Brice Bathellier, Lyubov Ushakova, and Simon Rumpel. Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron*, 76(2):435–449, 2012.

[20] Bruce P Bean. The action potential in mammalian central neurons. *Nature Reviews Neuroscience*, 8(6):451–465, 2007.

[21] Mark Bear, Barry Connors, and Michael A Paradiso. *Neuroscience: Exploring the brain*. Jones & Bartlett Learning, LLC, 2020.

[22] Mikhail Belkin. Fit without fear: remarkable mathematical phenomena of deep learning through the prism of interpolation. *Acta Numerica*, 30:203–248, 2021.

[23] Dirk Beutner, Thomas Voets, Erwin Neher, and Tobias Moser. Calcium dependence of exocytosis and endocytosis at the cochlear inner hair cell afferent synapse. *Neuron*, 29(3):681–690, 2001.

[24] Guo-qiang Bi and Mu-ming Poo. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18(24):10464–10472, 1998.

[25] Guo-qiang Bi and Mu-ming Poo. Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual review of neuroscience*, 24(1):139–166, 2001.

[26] Thomas Biederer, Pascal S Kaeser, and Thomas A Blanpied. Transcellular nanoalignment of synaptic function. *Neuron*, 96(3):680–696, 2017.

[27] Maria Blatow, Antonio Caputi, Nail Burnashev, Hannah Monyer, and Andrei Rozov. Ca2+ buffer saturation underlies paired pulse facilitation in calbindin-d28k-containing terminals. *Neuron*, 38(1):79–88, 2003.

[28] Martin Boerlin, Christian K Machens, and Sophie Denève. Predictive coding of dynamical variables in balanced spiking networks. *PLoS Computational Biology*, 9(11):e1003258, 2013.

[29] Johann H Bollmann, Bert Sakmann, and J Gerard G Borst. Calcium sensitivity of glutamate release in a calyx-type terminal. *Science*, 289(5481):953–957, 2000.

[30] Anna M Borghi, Ferdinand Binkofski, Cristiano Castelfranchi, Felice Cimatti, Claudia Scorolli, and Luca Tummolini. The challenge of abstract concepts. *Psychological Bulletin*, 143(3):263, 2017.

[31] J Gerard G Borst and John Soria van Hoeve. The calyx of held synapse: from model synapse to auditory relay. *Annual review of physiology*, 74:199–224, 2012.

[32] Matthew Botvinick, Jane X Wang, Will Dabney, Kevin J Miller, and Zeb Kurth-Nelson. Deep reinforcement learning and its neuroscientific implications. *Neuron*, 2020.

[33] Tiago Branco and Kevin Staras. The probability of neurotransmitter release: variability and feedback control at single synapses. *Nature Reviews Neuroscience*, 10(5):373–383, 2009.

[34] Braden AW Brinkman. Non-perturbative renormalization group analysis of nonlinear spiking networks. *arXiv preprint arXiv:2301.09600*, 2023.

[35] Axel T Brunger, Ucheor B Choi, Ying Lai, Jeremy Leitz, and Qiangjun Zhou. Molecular mechanisms of fast neurotransmitter release. *Annual review of biophysics*, 47:469–497, 2018.

[36] Zuzanna Brzosko, Susanna B Mierau, and Ole Paulsen. Neuromodulation of spike-timing-dependent plasticity: past, present, and future. *Neuron*, 103(4):563–581, 2019.

[37] Christian Büchel, Stephan Geuter, Christian Sprenger, and Falk Eippert. Placebo analgesia: a predictive coding perspective. *Neuron*, 81(6):1223–1239, 2014.

[38] Chong Chen, Rachel Satterfield, Samuel M Young Jr, and Peter Jonas. Triple function of synaptotagmin 7 ensures efficiency of high-frequency transmission at central gabaergic synapses. *Cell reports*, 21(8):2082–2089, 2017.

[39] In Ha Cho, Lauren C Panzera, Morven Chin, Scott A Alpizar, Genaro E Olveda, Robert A Hill, and Michael B Hoppa. The potassium channel subunit kvβ1 serves as a major control point for synaptic facilitation. *Proceedings of the National Academy of Sciences*, 117(47):29937–29947, 2020.

[40] Mark M Churchland, M Yu Byron, John P Cunningham, Leo P Sugrue, Marlene R Cohen, Greg S Corrado, William T Newsome, Andrew M Clark, Paymon Hosseini, Benjamin B Scott, et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature Neuroscience*, 13(3):369–378, 2010.

[41] R Cohen and D Atlas. R-type voltage-gated ca2+ channel interacts with synaptic proteins and recruits synaptotagmin to the plasma membrane of xenopus oocytes. *Neuroscience*, 128(4):831–841, 2004.

[42] Lee Cossell, Maria Florencia Iacaruso, Dylan R Muir, Rachael Houlton, Elie N Sader, Ho Ko, Sonja B Hofer, and Thomas D Mrsic-Flogel. Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*, 518(7539):399–403, 2015.

[43] Caitlin Mullan Crain, Kristy Kroeker, and Benjamin S Halpern. Interactive and cumulative effects of multiple human stressors in marine systems. *Ecology Letters*, 11(12):1304–1315, 2008.

[44] Alex Davies, Petar Veličković, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev, Richard Tanburn, Peter Battaglia, Charles Blundell, András Juhász, et al. Advancing mathematics by guiding human intuition with ai. *Nature*, 600(7887):70–74, 2021.

[45] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. 2005.

[46] Sophie Denève and Christian K Machens. Efficient codes and balanced networks. *Nature Neuroscience*, 19(3):375–382, 2016.

[47] Emily Jane Dennis, Ahmed El Hady, Angie Michaiel, Ann Clemens, Dougal R Gowan Tervo, Jakob Voigts, and Sandeep Robert Datta. Systems neuroscience of natural behaviors in rodents. *Journal of Neuroscience*, 41(5):911–919, 2021.

[48] Alain Destexhe, Zachary F Mainen, and Terrence J Sejnowski. Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism. *Journal of computational neuroscience*, 1(3):195–230, 1994.

[49] Ronald DeVore, Boris Hanin, and Guergana Petrova. Neural network approximation. *Acta Numerica*, 30:327–444, 2021.

[50] Jiajie Diao, Patricia Grob, Daniel J Cipriano, Minjoung Kyoung, Yunxiang Zhang, Sachi Shah, Amie Nguyen, Mark Padolina, Ankita Srivastava, Marija Vrljic, et al. Synaptic proteins promote calcium-triggered fast transition from point contact to full fusion. *elife*, 1:e00109, 2012.

[51] Lynn E Dobrunz and Charles F Stevens. Heterogeneity of release probability, facilitation, and depletion at central synapses. *Neuron*, 18(6):995–1008, 1997.

[52] FA Dodge Jr and Rami Rahamimoff. Co-operative action of calcium ions in transmitter release at the neuromuscular junction. *The Journal of physiology*, 193(2):419–432, 1967.

[53] Annette C Dolphin and Amy Lee. Presynaptic calcium channels: specialized control of synaptic neurotransmitter release. *Nature Reviews Neuroscience*, 21(4):213–229, 2020.

[54] Olga K Dudko, Gerhard Hummer, and Attila Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Physical review letters*, 96(10):108101, 2006.

[55] Gabriel Duncan, Katalin Rabl, Ian Gemp, Ruth Heidelberger, and Wallace B Thoreson. Quantitative analysis of synaptic release at the photoreceptor synapse. *Biophysical journal*, 98(10):2102–2110, 2010.

[56] Sami El-Boustani, Jacque PK Ip, Vincent Breton-Provencher, Graham W Knott, Hiroyuki Okuno, Haruhiko Bito, and Mriganka Sur. Locally coordinated synaptic plasticity of visual cortex neurons in vivo. *Science*, 360(6395):1349–1354, 2018.

[57] Steven J Eliades and Xiaoqin Wang. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*, 453(7198):1102–1106, 2008.

[58] Tamir Eliav, Shir R Maimon, Johnatan Aljadeff, Misha Tsodyks, Gily Ginosar, Liora Las, and Nachum Ulanovsky. Multiscale representation of very large environments in the hippocampus of flying bats. *Science*, 372(6545):eabg4020, 2021.

[59] Iddo Eliazar and Joseph Klafter. On the nonlinear modeling of shot noise. *Proceedings of the National Academy of Sciences*, 102(39):13779–13782, 2005.

[60] Daniel E Feldman. The spike-timing dependence of plasticity. *Neuron*, 75(4):556–571, 2012.

[61] Willliam Feller. *An introduction to probability theory and its applications, vol 2*. John Wiley & Sons, 2008.

[62] Rachel E Field, James A D'amour, Robin Tremblay, Christoph Miehl, Bernardo Rudy, Julijana Gjorgjieva, and Robert C Froemke. Heterosynaptic plasticity determines the set point for cortical excitatory-inhibitory balance. *Neuron*, 106(5):842–854, 2020.

[63] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, 2009.

[64] Peter Földiak. Forming sparse representations by local anti-hebbian learning. *Biological cybernetics*, 64(2):165–170, 1990.

[65] Galit Fuhrmann, Idan Segev, Henry Markram, and Misha Tsodyks. Coding of temporal information by activity-dependent synapses. *Journal of neurophysiology*, 2002.

[66] Ryota Fukaya, Marta Maglione, Stephan J Sigrist, and Takeshi Sakaba. Rapid ca2+ channel accumulation contributes to camp-mediated increase in transmission at hippocampal mossy fiber synapses. *Proceedings of the National Academy of Sciences*, 118(9), 2021.

[67] Shohei Furutachi, Alexis D Franklin, Thomas D Mrsic-Flogel, and Sonja B Hofer. Cooperative thalamocortical circuit mechanism for sensory prediction errors. *bioRxiv*, pages 2023–07, 2023.

[68] Stefano Fusi, Earl K Miller, and Mattia Rigotti. Why neurons mix: high dimensionality for higher cognition. *Current Opinion in Neurobiology*, 37:66–74, 2016.

[69] Ying Gao, Sylvain Zorman, Gregory Gundersen, Zhiqun Xi, Lu Ma, George Sirinakis, James E Rothman, and Yongli Zhang. Single reconstituted neuronal snare complexes zipper in three distinct stages. *Science*, 337(6100):1340–1343, 2012.

[70] Aleena R Garner and Georg B Keller. A cortical circuit for audio-visual predictions. *Nature Neuroscience*, 25(1):98–105, 2022.

[71] Tristan Geiller, Sadra Sadeh, Sebastian V Rolotti, Heike Blockus, Bert Vancura, Adrian Negrean, Andrew J Murray, Balázs Rózsa, Franck Polleux, Claudia Clopath, et al. Local circuit amplification of spatial selectivity in the hippocampus. *Nature*, pages 1–5, 2021.

[72] Apostolos P Georgopoulos, Andrew B Schwartz, and Ronald E Kettner. Neuronal population coding of movement direction. *Science*, 233(4771):1416–1419, 1986.

[73] Nicolas Gillis. *Nonnegative Matrix Factorization*. SIAM, 2020.

[74] Preeti Gipson, Yoshiyuki Fukuda, Radostin Danev, Ying Lai, Dong-Hua Chen, Wolfgang Baumeister, and Axel T Brunger. Morphologies of synaptic protein membrane fusion interfaces. *Proceedings of the National Academy of Sciences*, 114(34):9110–9115, 2017.

[75] Julijana Gjorgjieva, Claudia Clopath, Juliette Audet, and Jean-Pascal Pfister. A triplet spike-timing–dependent plasticity model generalizes the bienenstock–cooper–munro rule to higher-order spatiotemporal correlations. *Proceedings of the National Academy of Sciences*, 108(48):19383–19388, 2011.

[76] Maria E Gómez-Casati and Juan D Goutman. Divide and conquer acoustic diversity. *The EMBO Journal*, 40(5):e107531, 2021.

[77] Michael W Gramlich and Vitaly A Klyachko. Nanoscale organization of vesicle release at central synapses. *Trends in neurosciences*, 42(6):425–437, 2019.

[78] Michael Graupner and Nicolas Brunel. Mechanisms of induction and maintenance of spike-timing dependent plasticity in biophysical synapse models. *Frontiers in computational neuroscience*, 4:136, 2010.

[79] Michael Graupner and Nicolas Brunel. Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *Proceedings of the National Academy of Sciences*, 109(10):3991–3996, 2012.

[80] Kirill Grushin, Jing Wang, Jeff Coleman, James E Rothman, Charles V Sindelar, and Shyam S Krishnakumar. Structural basis for the clamping and ca 2+ activation of snare-mediated fusion by synaptotagmin. *Nature communications*, 10(1):1–12, 2019.

[81] Tatsuya Haga and Tomoki Fukai. Extended temporal association memory by modulations of inhibitory circuits. *Physical Review Letters*, 123(7):078101, 2019.

[82] A Pejmun Haghighi, Brian D McCabe, Richard D Fetter, Jessica E Palmer, Sabrina Hom, and Corey S Goodman. Retrograde control of synaptic transmission by postsynaptic camkii at the drosophila neuromuscular junction. *Neuron*, 39(2):255–267, 2003.

[83] Kristen M Harris and Richard J Weinberg. Ultrastructure of synapses in the mammalian brain. *Cold Spring Harbor perspectives in biology*, 4(5):a005587, 2012.

[84] Stephen C Harrison. Pictures of the prologue to neurotransmitter release. *Proceedings of the National Academy of Sciences*, 114(34):8920–8922, 2017.

[85] Ruth Heidelberger, Christian Heinemann, Erwin Neher, and Gary Matthews. Calcium dependence of the rate of exocytosis in a synaptic terminal. *Nature*, 371(6497):513–515, 1994.

[86] Itai Hershenhoren, Nevo Taaseh, Flora M Antunes, and Israel Nelken. Intracellular correlates of stimulus-specific adaptation. *Journal of Neuroscience*, 34(9):3303–3319, 2014.

[87] Loreen Hertäg and Claudia Clopath. Prediction-error neurons in circuits with multiple neuron types: Formation, refinement, and functional implications. *Proceedings of the National Academy of Sciences*, 119(13):e2115699119, 2022.

[88] Loreen Hertäg and Henning Sprekeler. Learning prediction error neurons in a canonical interneuron circuit. *Elife*, 9:e57541, 2020.

[89] David Higgins, Michael Graupner, and Nicolas Brunel. Memory maintenance in synapses with calcium-based plasticity in the presence of background activity. *PLoS Computational Biology*, 10(10):e1003834, 2014.

[90] Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500–544, 1952.

[91] Simone Holler, German Köstinger, Kevan AC Martin, Gregor FP Schuhknecht, and Ken J Stratford. Structure and function of a neocortical synapse. *Nature*, 591(7848):111–116, 2021.

[92] John F Houde and Michael I Jordan. Sensorimotor adaptation in speech production. *Science*, 279(5354):1213–1216, 1998.

[93] Yanping Huang and Rajesh PN Rao. Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5):580–593, 2011.

[94] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574, 1959.

[95] Enfu Hui, Jihong Bai, Ping Wang, Mutsuyuki Sugimori, Rodolfo R Llinas, and Edwin R Chapman. Three distinct kinetic groupings of the synaptotagmin family: candidate sensors for rapid and delayed exocytosis. *Proceedings of the National Academy of Sciences*, 102(14):5210–5214, 2005.

[96] Leena A Ibrahim, Lukas Mesik, Xu-ying Ji, Qi Fang, Hai-fu Li, Ya-tang Li, Brian Zingg, Li I Zhang, and Huizhong Whit Tao. Cross-modality sharpening of visual cortical processing through layer-1-mediated inhibition and disinhibition. *Neuron*, 89(5):1031–1045, 2016.

[97] Yanis Inglebert, Johnatan Aljadeff, Nicolas Brunel, and Dominique Debanne. Synaptic plasticity rules with physiological calcium levels. *Proceedings of the National Academy of Sciences*, 117(52):33639–33648, 2020.

[98] Giuliano Iurilli, Diego Ghezzi, Umberto Olcese, Glenda Lassi, Cristiano Nazzaro, Raffaella Tonini, Valter Tucci, Fabio Benfenati, and Paolo Medini. Sound-driven synaptic inhibition in primary visual cortex. *Neuron*, 73(4):814–828, 2012.

[99] Skyler L Jackman and Wade G Regehr. The mechanisms and functions of synaptic facilitation. *Neuron*, 94(3):447–464, 2017.

[100] Skyler L Jackman, Josef Turecek, Justine E Belinsky, and Wade G Regehr. The calcium sensor synaptotagmin 7 is required for synaptic facilitation. *Nature*, 529(7584):88–91, 2016.

[101] Stuart L Johnson, Christoph Franz, Stephanie Kuhn, David N Furness, Lukas Rüttiger, Stefan Münkner, Marcelo N Rivolta, Elizabeth P Seward, Harvey R Herschman, Jutta Engel, et al. Synaptotagmin iv determines the linear ca 2+ dependence of vesicle fusion at auditory ribbon synapses. *Nature neuroscience*, 13(1):45–52, 2010.

[102] Rebecca Jordan and Georg B Keller. Opposing influence of top-down and bottom-up input on excitatory layer 2/3 neurons in mouse primary visual cortex. *Neuron*, 108(6):1194–1206, 2020.

[103] Jonathan Kadmon, Jonathan Timcheck, and Surya Ganguli. Predictive coding in balanced neural networks with noise, chaos and delays. *Advances in Neural Information Processing Systems*, 33:16677–16688, 2020.

[104] Pascal S Kaeser and Wade G Regehr. Molecular mechanisms for synchronous, asynchronous, and spontaneous neurotransmitter release. *Annual review of physiology*, 76:333–363, 2014.

[105] Pascal S Kaeser and Wade G Regehr. The readily releasable pool of synaptic vesicles. *Current opinion in neurobiology*, 43:63–70, 2017.

[106] Eric R Kandel. The molecular biology of memory storage: a dialogue between genes and synapses. *Science*, 294(5544):1030–1038, 2001.

[107] Eric R Kandel and James H Schwartz. Molecular biology of learning: modulation of transmitter release. *Science*, 218(4571):433–443, 1982.

[108] Haruo Kasai, Noriko Takahashi, and Hiroshi Tokumaru. Distinct initial snare configurations underlying the diversity of exocytosis. *Physiological reviews*, 92(4):1915–1964, 2012.

[109] Bernard Katz and Ricardo Miledi. The measurement of synaptic delay, and the time course of acetylcholine release at the neuromuscular junction. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 161(985):483–495, 1965.

[110] Shin-ya Kawaguchi and Takeshi Sakaba. Fast ca2+ buffer-dependent reliable but plastic transmission at small cns synapses revealed by direct bouton recording. *Cell reports*, 21(12):3338–3345, 2017.

[111] Tara Keck, Taro Toyoizumi, Lu Chen, Brent Doiron, Daniel E Feldman, Kevin Fox, Wulfram Gerstner, Philip G Haydon, Mark Hübener, Hey-Kyoung Lee, et al. Integrating hebbian and homeostatic plasticity: the current state of the field and future research directions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1715):20160158, 2017.

[112] Georg B Keller, Tobias Bonhoeffer, and Mark Hübener. Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron*, 74(5):809–815, 2012.

[113] Georg B Keller and Richard HR Hahnloser. Neural processing of auditory feedback during vocal practice in a songbird. *Nature*, 457(7226):187–190, 2009.

[114] Georg B Keller and Thomas D Mrsic-Flogel. Predictive processing: a canonical cortical computation. *Neuron*, 100(2):424–435, 2018.

[115] Richard Kempter, Wulfram Gerstner, and J Leo van Hemmen. Hebbian learning and spiking neurons. *Physical Review E*, 59(4):4498, 1999.

[116] Angharad M Kerr, Ellen Reisinger, and Peter Jonas. Differential dependence of phasic transmitter release on synaptotagmin 1 at gabaergic and glutamatergic hippocampal synapses. *Proceedings of the National Academy of Sciences*, 105(40):15581–15586, 2008.

[117] Adil G Khan, Jasper Poort, Angus Chadwick, Antonin Blot, Maneesh Sahani, Thomas D Mrsic-Flogel, and Sonja B Hofer. Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. *Nature Neuroscience*, 21(6):851–859, 2018.

[118] Anmo J Kim, Lisa M Fenk, Cheng Lyu, and Gaby Maimon. Quantitative predictions orchestrate visual signaling in *drosophila*. *Cell*, 168(1):280–294, 2017.

[119] John Frank Charles Kingman. *Poisson processes*, volume 3. Clarendon Press, 1992.

[120] Jürgen Klingauf and Erwin Neher. Modeling buffered ca2+ diffusion near the membrane: implications for secretion in neuroendocrine cells. *Biophysical journal*, 72(2):674–690, 1997.

[121] Ho Ko, Lee Cossell, Chiara Baragli, Jan Antolik, Claudia Clopath, Sonja B Hofer, and Thomas D Mrsic-Flogel. The emergence of functional microcircuits in visual cortex. *Nature*, 496(7443):96–100, 2013.

[122] Ho Ko, Sonja B Hofer, Bruno Pichler, Katherine A Buchanan, P Jesper Sjöström, and Thomas D Mrsic-Flogel. Functional specificity of local synaptic connections in neocortical networks. *Nature*, 473(7345):87–91, 2011.

[123] Janus RL Kobbersmed, Andreas T Grasskamp, Meida Jusyte, Mathias A Böhme, Susanne Ditlevsen, Jakob Balslev Sørensen, and Alexander M Walter. Rapid regulation of vesicle priming explains synaptic facilitation despite heterogeneous vesicle: Ca2+ channel distances. *Elife*, 9:e51032, 2020.

[124] Olexiy Kochubey, Norbert Babai, and Ralf Schneggenburger. A synaptotagmin isoform switch during the development of an identified cns synapse. *Neuron*, 90(5):984–999, 2016.

[125] Olexiy Kochubey, Yunyun Han, and Ralf Schneggenburger. Developmental regulation of the intracellular ca2+ sensitivity of vesicle fusion and ca2+–secretion coupling at the rat calyx of held. *The Journal of physiology*, 587(12):3009–3023, 2009.

[126] Olexiy Kochubey, Xuelin Lou, and Ralf Schneggenburger. Regulation of transmitter release by ca2+ and synaptotagmin: insights from a large cns synapse. *Trends in neurosciences*, 34(5):237–246, 2011.

[127] Yaroslav Felipe Kalle Kossio, Sven Goedeke, Christian Klos, and Raoul-Martin Memmesheimer. Drifting assemblies for persistent memory: Neuron transitions and unsupervised compensation. *Proceedings of the National Academy of Sciences*, 118(46), 2021.

[128] Jeanette Hellgren Kotaleski and Kim T Blackwell. Modelling the molecular mechanisms of synaptic plasticity using systems biology approaches. *Nature Reviews Neuroscience*, 11(4):239–251, 2010.

[129] Hendrik Anthony Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[130] Dhireesha Kudithipudi, Mario Aguilar-Simon, Jonathan Babb, Maxim Bazhenov, Douglas Blackiston, Josh Bongard, Andrew P Brna, Suraj Chakravarthi Raja, Nick Cheney, Jeff Clune, et al. Biological underpinnings for lifelong learning machines. *Nature Machine Intelligence*, 4(3):196–210, 2022.

[131] Łukasz Kuśmierz, Shun Ogawa, and Taro Toyoizumi. Edge of chaos and avalanches in neural networks with heavy-tailed synaptic weight distribution. *Physical Review Letters*, 125:028101, Jul 2020.

[132] Grant F Kusick, Morven Chin, Sumana Raychaudhuri, Kristina Lippmann, Kadidia P Adula, Edward J Hujber, Thien Vu, M Wayne Davis, Erik M Jorgensen, and Shigeki Watanabe. Synaptic vesicles transiently dock to refill release sites. *Nature neuroscience*, 23(11):1329–1338, 2020.

[133] Dae-Hyuk Kweon, Byoungjae Kong, and Yeon-Kyun Shin. Hemifusion in synaptic vesicle cycle. *Frontiers in Molecular Neuroscience*, 10:65, 2017.

[134] Minjoung Kyoung, Ankita Srivastava, Yunxiang Zhang, Jiajie Diao, Marija Vrljic, Patricia Grob, Eva Nogales, Steven Chu, and Axel T Brunger. In vitro system capable of differentiating fast ca^{2+}-triggered content mixing from lipid exchange for mechanistic studies of neurotransmitter release. *Proceedings of the National Academy of Sciences*, 108(29):E304–E313, 2011.

[135] Guillaume Lajoie, Nedialko I Krouchev, John F Kalaska, Adrienne L Fairhall, and Eberhard E Fetz. Correlation-based model of artificially induced plasticity in motor cortex by a bidirectional brain-computer interface. *PLoS Computational Biology*, 13(2):e1005343, 2017.

[136] Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. Building machines that learn and think like people. *Behavioral and brain sciences*, 40:e253, 2017.

[137] George Lakoff and Rafael Núñez. *Where mathematics comes from*, volume 6. New York: Basic Books, 2000.

[138] Christopher Langdon, Mikhail Genkin, and Tatiana A Engel. A unifying perspective on neural manifolds and circuits for cognition. *Nature Reviews Neuroscience*, pages 1–15, 2023.

[139] Rylan S Larsen and P Jesper Sjöström. Synapse-type-specific plasticity in local circuits. *Current opinion in neurobiology*, 35:127–135, 2015.

[140] Yann LeCun. A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27. *Open Review*, 62, 2022.

[141] Marcus Leinweber, Daniel R Ward, Jan M Sobczak, Alexander Attinger, and Georg B Keller. A sensorimotor circuit in mouse cortex for visual flow predictions. *Neuron*, 95(6):1420–1432, 2017.

[142] Ashok Litwin-Kumar and Brent Doiron. Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature Communications*, 5(1):1–12, 2014.

[143] Yoav Livneh and Mark L Andermann. Cellular activity in insular cortex across seconds to hours: Sensations and predictions of bodily states. *Neuron*, 109(22):3576–3593, 2021.

[144] Yoav Livneh, Arthur U Sugden, Joseph C Madara, Rachel A Essner, Vanessa I Flores, Lauren A Sugden, Jon M Resch, Bradford B Lowell, and Mark L Andermann. Estimation of current and future physiological states in insular cortex. *Neuron*, 105(6):1094–1111, 2020.

[145] Xuelin Lou, Volker Scheuss, and Ralf Schneggenburger. Allosteric modulation of the presynaptic ca^{2+} sensor for vesicle fusion. *Nature*, 435(7041):497, 2005.

[146] Fujun Luo, Taulant Bacaj, and Thomas C Südhof. Synaptotagmin-7 is essential for ca2+-triggered delayed asynchronous release but not for ca2+-dependent vesicle priming in retinal ribbon synapses. *Journal of Neuroscience*, 35(31):11024–11033, 2015.

[147] Fujun Luo and Thomas C Südhof. Synaptotagmin-7-mediated asynchronous release boosts high-fidelity synchronous transmission at a central synapse. *Neuron*, 94(4):826–839, 2017.

[148] KL Magleby. The effect of repetitive stimulation on facilitation of transmitter release at the frog neuromuscular junction. *The Journal of physiology*, 234(2):327–352, 1973.

[149] Marcelo O Magnasco, Oreste Piro, and Guillermo A Cecchi. Self-tuned critical anti-hebbian networks. *Physical Review Letters*, 102(25):258102, 2009.

[150] Tuomo Mäki-Marttunen, Nicolangelo Iannella, Andrew G Edwards, Gaute T Einevoll, and Kim T Blackwell. A unified computational model for cortical post-synaptic plasticity. *Elife*, 9:e55714, 2020.

[151] Hiroshi Makino and Takaki Komiyama. Learning enhances the relative impact of top-down processing in the visual cortex. *Nature Neuroscience*, 18(8):1116–1122, 2015.

[152] Gerardo Malagon, Takafumi Miki, Isabel Llano, Erwin Neher, and Alain Marty. Counting vesicular release events reveals binomial release statistics at single glutamatergic synapses. *Journal of Neuroscience*, 36(14):4010–4025, 2016.

[153] Robert C Malenka and Mark F Bear. Ltp and ltd: an embarrassment of riches. *Neuron*, 44(1):5–21, 2004.

[154] Fabio Manca, Frederic Pincet, Lev Truskinovsky, James E Rothman, Lionel Foret, and Matthieu Caruel. Snare machinery is optimized for ultrafast fusion. *Proceedings of the National Academy of Sciences*, 116(7):2435–2442, 2019.

[155] Alon Manor and Nadav M Shnerb. Multiplicative noise and second order phase transitions. *Physical Review Letters*, 103(3):030601, 2009.

[156] Valerio Mante, David Sussillo, Krishna V Shenoy, and William T Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474):78–84, 2013.

[157] Eve Marder. Neuromodulation of neuronal circuits: back to the future. *Neuron*, 76(1):1–11, 2012.

[158] Bruno Marks and Harvey T McMahon. Calcium triggers calcineurin-dependent synaptic vesicle recycling in mammalian nerve terminals. *Current biology*, 8(13):740–749, 1998.

[159] Mark Mayford, Steven A Siegelbaum, and Eric R Kandel. Synapses and memory storage. *Cold Spring Harbor perspectives in biology*, 4(6):a005751, 2012.

[160] Takafumi Miki, Walter A Kaufmann, Gerardo Malagon, Laura Gomez, Katsuhiko Tabuchi, Masahiko Watanabe, Ryuichi Shigemoto, and Alain Marty. Numbers of presynaptic ca2+ channel clusters match those of functionally defined vesicular docking sites in single central synapses. *Proceedings of the National Academy of Sciences*, 114(26):E5246–E5255, 2017.

[161] Takafumi Miki, Yukihiro Nakamura, Gerardo Malagon, Erwin Neher, and Alain Marty. Two-component latency distributions indicate two-step vesicular release at simple glutamatergic synapses. *Nature communications*, 9(1):3943, 2018.

[162] Fabian A Mikulasch, Lucas Rudelt, and Viola Priesemann. Local dendritic balance enables learning of efficient representations in networks of spiking neurons. *Proceedings of the National Academy of Sciences*, 118(50):e2021925118, 2021.

[163] Fabian A Mikulasch, Lucas Rudelt, Michael Wibral, and Viola Priesemann. Where is the error? hierarchical predictive coding through dendritic error computation. *Trends in Neurosciences*, 46(1):45–59, 2023.

[164] Ralf Mohrmann, Heidi De Wit, Matthijs Verhage, Erwin Neher, and Jakob B Sørensen. Fast vesicle fusion in living cells requires at least three snare complexes. *Science*, 330(6003):502–505, 2010.

[165] Hakhamanesh Mostafavi, Sathish Thiyagarajan, Benjamin S Stratton, Erdem Karatekin, Jason M Warner, James E Rothman, and Ben O'Shaughnessy. Entropic forces drive self-organization and membrane fusion by snare proteins. *Proceedings of the National Academy of Sciences*, 114(21):5455–5460, 2017.

[166] Martin Müller, Felix Felmy, Beat Schwaller, and Ralf Schneggenburger. Parvalbumin is a mobile presynaptic ca2+ buffer in the calyx of held that accelerates the decay of ca2+ and short-term facilitation. *Journal of Neuroscience*, 27(9):2261–2271, 2007.

[167] Yukihiro Nakamura, Harumi Harada, Naomi Kamasawa, Ko Matsui, Jason S Rothman, Ryuichi Shigemoto, R Angus Silver, David A DiGregorio, and Tomoyuki Takahashi. Nanoscale distribution of presynaptic ca2+ channels and its impact on vesicular release during development. *Neuron*, 85(1):145–158, 2015.

[168] Evanthia Nanou, Todd Scheuer, and William A Catterall. Calcium sensor regulation of the cav2. 1 ca2+ channel contributes to long-term potentiation and spatial learning. *Proceedings of the National Academy of Sciences*, 113(46):13209–13214, 2016.

[169] Erwin Neher. Usefulness and limitations of linear approximations to the understanding of ca++ signals. *Cell calcium*, 24(5-6):345–357, 1998.

[170] Erwin Neher. What is rate-limiting during sustained synaptic activity: vesicle supply or the availability of release sites. *Frontiers in synaptic neuroscience*, 2:144, 2010.

[171] Erwin Neher. Merits and limitations of vesicle pool models in view of heterogeneous populations of synaptic vesicles. *Neuron*, 87(6):1131–1142, 2015.

[172] Erwin Neher. Some subtle lessons from the calyx of held synapse. *Biophysical Journal*, 112(2):215–223, 2017.

[173] Erwin Neher and Nils Brose. Dynamically primed synaptic vesicle states: key to understand synaptic short-term plasticity. *Neuron*, 100(6):1283–1291, 2018.

[174] Anders Nelson, David M Schneider, Jun Takatoh, Katsuyasu Sakurai, Fan Wang, and Richard Mooney. A circuit for motor cortical modulation of auditory cortical activity. *Journal of Neuroscience*, 33(36):14342–14353, 2013.

[175] Sacha B Nelson. Synaptic plasticity: taming the beast. *Nature neuroscience*, 3(11):1178–1183, 2000.

[176] Thomas Nevian and Bert Sakmann. Spine ca^{2+} signaling in spike-timing-dependent plasticity. *Journal of Neuroscience*, 26(43):11001–11013, 2006.

[177] Makoto Nishiyama, Kyonsoo Hong, Katsuhiko Mikoshiba, Mu-Ming Poo, and Kunio Kato. Calcium stores regulate the polarity and input specificity of synaptic modification. *Nature*, 408(6812):584–588, 2000.

[178] Gabriel Koch Ocker. Dynamics of stochastic integrate-and-fire networks. *Physical Review X*, 12(4):041007, 2022.

[179] Gabriel Koch Ocker, Ashok Litwin-Kumar, and Brent Doiron. Self-organization of microcircuits in networks of spiking neurons with plastic synapses. *PLoS Computational Biology*, 11(8):e1004458, 2015.

[180] Marieelen Oelkers, Hannes Witt, Partho Halder, Reinhard Jahn, and Andreas Janshoff. Snare-mediated membrane fusion trajectories derived from force-clamp experiments. *Proceedings of the National Academy of Sciences*, 113(46):13051–13056, 2016.

[181] John O'Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 1971.

[182] Srdjan Ostojic and Stefano Fusi. The computational role of structure in neural activity and connectivity. *arXiv preprint arXiv:2308.16772*, 2023.

[183] Özge D Özçete and Tobias Moser. A sensory cell diversifies its output by varying ca2+ influx-release coupling among active zones. *The EMBO journal*, 40(5):e106010, 2021.

[184] Sean M O'Toole, Hassana K Oyibo, and Georg B Keller. Molecularly targetable cell types in mouse visual cortex have distinguishable prediction error responses. *Neuron*, 2023.

[185] Bin Pan and Robert S Zucker. A general model of synaptic transmission and short-term plasticity. *Neuron*, 62(4):539–554, 2009.

[186] Giovanni Peccati and Matthias Reitzner. *Stochastic analysis for Poisson point processes: Malliavin calculus, Wiener-Itô chaos expansions and stochastic geometry*, volume 7. Springer, 2016.

[187] Ulises Pereira and Nicolas Brunel. Attractor dynamics in networks with learning rules inferred from in vivo data. *Neuron*, 99(1):227–238, 2018.

[188] Ulises Pereira-Obilinovic, Johnatan Aljadeff, and Nicolas Brunel. Forgetting leads to chaos in attractor networks. *Physical Review X*, 13(1):011009, 2023.

[189] Edward Perez-Reyes. Molecular physiology of low-voltage-activated t-type calcium channels. *Physiological reviews*, 83(1):117–161, 2003.

[190] Jasper Poort, Adil G Khan, Marius Pachitariu, Abdellatif Nemri, Ivana Orsolic, Julija Krupic, Marius Bauza, Maneesh Sahani, Georg B Keller, Thomas D Mrsic-Flogel, et al. Learning enhances sensory and multiple non-sensory representations in primary visual cortex. *Neuron*, 86(6):1478–1490, 2015.

[191] Abhijith Radhakrishnan, Xia Li, Kirill Grushin, Shyam S Krishnakumar, Jun Liu, and James E Rothman. Symmetrical arrangement of proteins under release-ready vesicles in presynaptic terminals. *Proceedings of the National Academy of Sciences*, 118(5), 2021.

[192] Anand Radhakrishnan, Alexander Stein, Reinhard Jahn, and Dirk Fasshauer. The ca2+ affinity of synaptotagmin 1 is markedly increased by a specific interaction of its c2b domain with phosphatidylinositol 4, 5-bisphosphate. *Journal of Biological Chemistry*, 284(38):25749–25760, 2009.

[193] Rajesh PN Rao and Dana H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999.

[194] Neta Ravid Tannenbaum and Yoram Burak. Shaping neural circuits by high order synaptic interactions. *PLoS Computational Biology*, 12(8):e1005056, 2016.

[195] Nelson Rebola, Maria Reva, Tekla Kirizs, Miklos Szoboszlay, Andrea Lőrincz, Gael Moneron, Zoltan Nusser, and David A DiGregorio. Distinct nanoscale calcium channel and synaptic vesicle topographies contribute to the diversity of synaptic function. *Neuron*, 104(4):693–710, 2019.

[196] Wade G Regehr. Short-term presynaptic plasticity. *Cold Spring Harbor perspectives in biology*, 4(7):a005702, 2012.

[197] Magnus J E Richardson and Rupert Swarbrick. Firing-rate response of a neuron receiving excitatory and inhibitory synaptic shot noise. *Physical Review Letters*, 105(17):178102, 2010.

[198] Mattia Rigotti, Omri Barak, Melissa R Warden, Xiao-Jing Wang, Nathaniel D Daw, Earl K Miller, and Stefano Fusi. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451):585–590, 2013.

[199] Robert Rosenbaum, Jonathan Rubin, and Brent Doiron. Short term synaptic depression imposes a frequency dependent filter on synaptic information transfer. *PLoS Comput Biol*, 8(6):e1002557, 2012.

[200] Christopher J Rozell, Don H Johnson, Richard G Baraniuk, and Bruno A Olshausen. Sparse coding via thresholding and local competition in neural circuits. *Neural computation*, 20(10):2526–2563, 2008.

[201] Andrei Rozov, Alexey P Bolshakov, and Fliza Valiullina-Rakhmatullina. The ever-growing puzzle of asynchronous release. *Frontiers in cellular neuroscience*, 13:28, 2019.

[202] Andrej Rozov, Nail Burnashev, Bert Sakmann, and Erwin Neher. Transmitter release modulation by intracellular ca2+ buffers in facilitating and depressing nerve terminals of pyramidal cells in layer 2/3 of the rat neocortex indicates a target cell-specific difference in presynaptic calcium dynamics. *The Journal of physiology*, 531(3):807–826, 2001.

[203] Jonathan Rubin, Daniel D Lee, and H Sompolinsky. Equilibrium properties of temporally asymmetric hebbian plasticity. *Physical Review Letters*, 86(2):364, 2001.

[204] Katy J Rubin, Gunnar Pruessner, and Grigorios A Pavliotis. Mapping multiplicative to additive noise. *Journal of Physics A: Mathematical and Theoretical*, 47(19):195001, 2014.

[205] Takeshi Sakaba. Two ca2+-dependent steps controlling synaptic vesicle fusion and replenishment at the cerebellar basket cell terminal. *Neuron*, 57(3):406–419, 2008.

[206] Raphaël Sarfati, Julie C Hayes, Élie Sarfati, and Orit Peleg. Spatio-temporal reconstruction of emergent flash synchronization in firefly swarms via stereoscopic 360-degree cameras. *Journal of the Royal Society Interface*, 17(170):20200179, 2020.

[207] Michael Schaefer, Michael Sommer, and Martin Karplus. ph-dependence of protein stability: absolute electrostatic free energy differences between conformations. *The Journal of Physical Chemistry B*, 101(9):1663–1683, 1997.

[208] Ralf Schneggenburger and Erwin Neher. Intracellular calcium dependence of transmitter release rates at a fast central synapse. *Nature*, 406(6798):889, 2000.

[209] David M Schneider, Anders Nelson, and Richard Mooney. A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*, 513(7517):189–194, 2014.

[210] David M Schneider, Janani Sundararajan, and Richard Mooney. A cortical filter that learns to suppress the acoustic consequences of movement. *Nature*, 561(7723):391–395, 2018.

[211] Cornelius Schroeder, Jonathan Oesterle, Philipp Berens, Takeshi Yoshimatsu, and Tom Baden. Distinct synaptic transfer functions in same-type photoreceptors. *Elife*, 10:e67851, 2021.

[212] Wolfram Schultz. Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80(1):1–27, 1998.

[213] Beat Schwaller. Cytosolic ca2+ buffers are inherently ca2+ signal modulators. *Cold Spring Harbor Perspectives in Biology*, 12(1):a035543, 2020.

[214] Harel Z Shouval, Mark F Bear, and Leon N Cooper. A unified model of NMDA receptor-dependent bidirectional synaptic plasticity. *Proceedings of the National Academy of Sciences*, 99(16):10831–10836, 2002.

[215] Raunak Sinha, Saheeb Ahmed, Reinhard Jahn, and Jurgen Klingauf. Two synaptobrevin molecules are sufficient for vesicle fusion in central nervous system synapses. *Proceedings of the National Academy of Sciences*, 108(34):14318–14323, 2011.

[216] Donald L Snyder and Michael I Miller. *Random point processes in time and space*. Springer Science & Business Media, 2012.

[217] Sen Song, Per Jesper Sjöström, Markus Reigl, Sacha Nelson, and Dmitri B Chklovskii. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology*, 3(3):e68, 2005.

[218] Daniel Soudry, Elad Hoffer, Mor Shpigel Nacson, Suriya Gunasekar, and Nathan Srebro. The implicit bias of gradient descent on separable data. *The Journal of Machine Learning Research*, 19(1):2822–2878, 2018.

[219] Michael W Spratling. Predictive coding as a model of biased competition in visual attention. *Vision Research*, 48(12):1391–1408, 2008.

[220] Mandyam Veerambudi Srinivasan, Simon Barry Laughlin, and Andreas Dubs. Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 216(1205):427–459, 1982.

[221] Elise F Stanley. The nanophysiology of fast transmitter release. *Trends in Neurosciences*, 39(3):183–197, 2016.

[222] Charles F Stevens. Models are common; good theories are scarce. *nature neuroscience*, 3(11):1177–1177, 2000.

[223] Thomas C Südhof. Calcium control of neurotransmitter release. *Cold Spring Harbor perspectives in biology*, 4(1):a011353, 2012.

[224] Thomas C Südhof. Neurotransmitter release: the last millisecond in the life of a synaptic vesicle. *Neuron*, 80(3):675–690, 2013.

[225] 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.

[226] Erica Tagliatti, Oscar D Bello, Philipe RF Mendonça, Dimitrios Kotzadimitriou, Elizabeth Nicholson, Jeff Coleman, Yulia Timofeeva, James E Rothman, Shyam S Krishnakumar, and Kirill E Volynski. Synaptotagmin 1 oligomers clamp and regulate different modes of neurotransmitter release. *Proceedings of the National Academy of Sciences*, 117(7):3819–3827, 2020.

[227] Hideki Takayasu, Aki-Hiro Sato, and Misako Takayasu. Stable infinite variance fluctuations in randomly amplified Langevin systems. *Physical Review Letters*, 79(6):966, 1997.

[228] David W Tank, WG Regehr, and KR Delaney. A quantitative analysis of presynaptic calcium dynamics that contribute to short-term enhancement. *Journal of Neuroscience*, 15(12):7940–7952, 1995.

[229] Bosiljka Tasic, Zizhen Yao, Lucas T Graybuck, Kimberly A Smith, Thuc Nghi Nguyen, Darren Bertagnolli, Jeff Goldy, Emma Garren, Michael N Economo, Sarada Viswanathan, et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, 563(7729):72–78, 2018.

[230] Misha V Tsodyks and Henry Markram. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the national academy of sciences*, 94(2):719–723, 1997.

[231] Josef Turecek, Skyler L Jackman, and Wade G Regehr. Synaptic specializations support frequency-independent purkinje cell output from the cerebellar cortex. *Cell reports*, 17(12):3256–3268, 2016.

[232] Josef Turecek and Wade G Regehr. Synaptotagmin 7 mediates both facilitation and asynchronous release at granule cell synapses. *Journal of Neuroscience*, 38(13):3240–3251, 2018.

[233] Nachum Ulanovsky, Liora Las, Dina Farkas, and Israel Nelken. Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience*, 24(46):10440–10453, 2004.

[234] Geert van den Bogaart and Reinhard Jahn. Counting the snares needed for membrane fusion. *Journal of molecular cell biology*, 3(4):204–205, 2011.

[235] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018.

[236] Thomas Voets. Dissection of three ca2+-dependent steps leading to secretion in chromaffin cells from mouse adrenal slices. *Neuron*, 28(2):537–545, 2000.

[237] Kirill E Volynski and Shyam S Krishnakumar. Synergistic control of neurotransmitter release by different members of the synaptotagmin family. *Current opinion in neurobiology*, 51:154–162, 2018.

[238] Nicholas P Vyleta and Peter Jonas. Loose coupling between ca2+ channels and release sensors at a plastic hippocampal synapse. *Science*, 343(6171):665–670, 2014.

[239] Kevin S Walsh, David P McGovern, Andy Clark, and Redmond G O'Connell. Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, 1464(1):242–268, 2020.

[240] Bin Wang and Johnatan Aljadeff. Multiplicative shot-noise: A new route to stability of plastic networks. *Physical Review Letters*, 129(6):068101, 2022.

[241] Bin Wang and Olga K Dudko. A theory of synaptic transmission. *Elife*, 10:e73585, 2021.

[242] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023.

[243] Jing Wang, Oscar Bello, Sarah M Auclair, Jeff Coleman, Frederic Pincet, Shyam S Krishnakumar, Charles V Sindelar, and James E Rothman. Calcium sensitive ring-like oligomers formed by synaptotagmin. *Proceedings of the National Academy of Sciences*, 111(38):13966–13971, 2014.

[244] Samuel S-H Wang, Winfried Denk, and Michael Häusser. Coincidence detection in single dendritic spines mediated by calcium release. *Nature Neuroscience*, 3(12):1266–1273, 2000.

[245] Felix C Widmer, Sean M O'Toole, and Georg B Keller. NMDA receptors in visual cortex are necessary for normal visuomotor integration and skill learning. *Elife*, 11:e71476, 2022.

[246] Thorsten Winkler, Ulrich Kettling, Andre Koltermann, and Manfred Eigen. Confocal fluorescence coincidence analysis: an approach to ultra high-throughput screening. *Proceedings of the National Academy of Sciences*, 96(4):1375–1378, 1999.

[247] Markus Wölfel, Xuelin Lou, and Ralf Schneggenburger. A mechanism intrinsic to the vesicle fusion machinery determines fast and slow transmitter release at a large cns synapse. *Journal of Neuroscience*, 27(12):3198–3210, 2007.

[248] Anne C Wolfes and Camin Dean. The diversity of synaptotagmin isoforms. *Current Opinion in Neurobiology*, 63:198–209, 2020.

[249] Theodosia Woo, Xitong Liang, Dominic A Evans, Olivier Fernandez, Friedrich Kretschmer, Sam Reiter, and Gilles Laurent. The dynamics of pattern matching in camouflaging cuttlefish. *Nature*, pages 1–7, 2023.

[250] Daniel LK Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3):356–365, 2016.

[251] Guangyu Robert Yang and Manuel Molano-Mazón. Towards the next generation of recurrent network models for cognitive neuroscience. *Current opinion in neurobiology*, 70:182–192, 2021.

[252] Yan Yang and Kevin D Gillis. A highly ca2+-sensitive pool of granules is regulated by glucose and protein kinases in insulin-secreting ins-1 cells. *The Journal of general physiology*, 124(6):641–651, 2004.

[253] Ying Yang and Nicole Calakos. Presynaptic long-term plasticity. *Frontiers in synaptic neuroscience*, 5:8, 2013.

[254] Tohar S Yarden, Adi Mizrahi, and Israel Nelken. Context-dependent inhibitory control of stimulus-specific adaptation. *Journal of Neuroscience*, 42(23):4629–4651, 2022.

[255] Friedemann Zenke, Wulfram Gerstner, and Surya Ganguli. The temporal paradox of hebbian learning and homeostatic plasticity. *Current Opinion in Neurobiology*, 43:166–176, 2017.

[256] Yaojun Zhang and Olga K Dudko. Statistical mechanics of viral entry. *Physical review letters*, 114(1):018104, 2015.

[257] Jie Zhu, Zachary A McDargh, Feng Li, Shyam Krishnakumar, James E Rothman, and Ben O'Shaughnessy. Synaptotagmin rings as high sensitivity regulators of synaptic vesicle docking and fusion. *bioRxiv*, 2021.

[258] Yaniv Ziv, Laurie D Burns, Eric D Cocker, Elizabeth O Hamel, Kunal K Ghosh, Lacey J Kitch, Abbas El Gamal, and Mark J Schnitzer. Long-term dynamics of ca1 hippocampal place codes. *Nature Neuroscience*, 16(3):264–266, 2013.

[259] Pawel Zmarz and Georg B Keller. Mismatch receptive fields in mouse visual cortex. *Neuron*, 92(4):766–772, 2016.

[260] Robert S Zucker. Exocytosis: a molecular and physiological perspective. *Neuron*, 17(6):1049–1055, 1996.