







TOPICAL REVIEW

How to incorporate biological insights into network models and why it matters

Laura Bernáez Timón¹ , Pierre Ekelmans² , Nataliya Kravnyukova³ , Tobias Rose³ ,
Laura Busse^{4,5}  and Tatjana Tchumatchenko^{1,3} 

¹Institute for Physiological Chemistry, University of Mainz Medical Center, Mainz, Germany

²Frankfurt Institute for Advanced Studies, Frankfurt, Germany

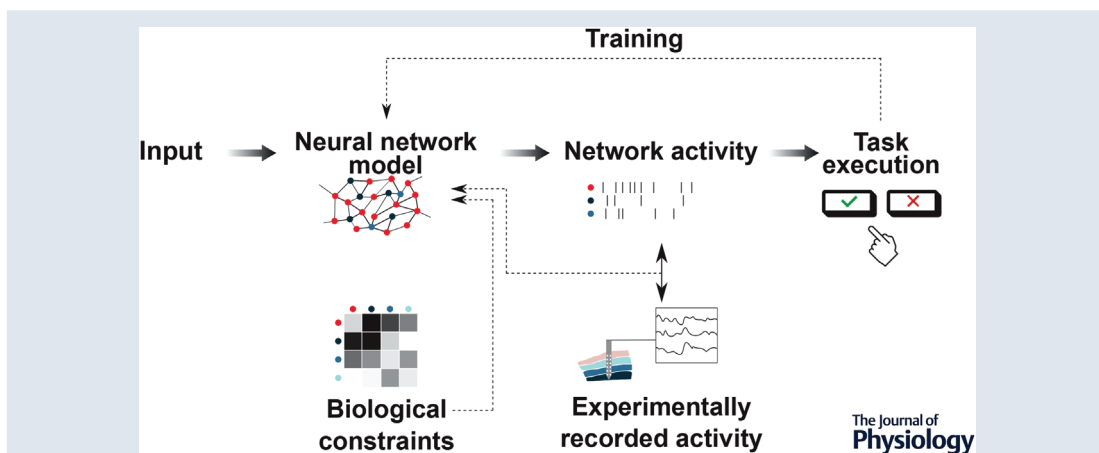
³Institute of Experimental Epileptology and Cognition Research, University of Bonn Medical Center, Bonn, Germany

⁴Division of Neurobiology, Faculty of Biology, LMU Munich, Munich, Germany

⁵Bernstein Center for Computational Neuroscience, Munich, Germany

Handling Editors: Katalin Toth & Michael Okun

The peer review history is available in the Supporting Information section of this article (<https://doi.org/10.1113/JP282755#support-information-section>).



Abstract Due to the staggering complexity of the brain and its neural circuitry, neuroscientists rely on the analysis of mathematical models to elucidate its function. From Hodgkin and Huxley's detailed description of the action potential in 1952 to today, new theories and increasing computational power have opened up novel avenues to study how neural circuits implement the computations that underlie behaviour. Computational neuroscientists have developed many models of neural circuits that differ in complexity, biological realism or

Laura Bernáez (PhD candidate), Pierre Ekelmans (PhD candidate) and Nataliya Kravnyukova (Post-doc) conduct research in the group led by Professor Tchumatchenko. They develop biologically grounded mathematical theories of neural networks in close collaborations with experimental laboratories, including the labs of Professor Tobias Rose and Professor Laura Busse. Laura, Pierre, and Nataliya believe that the boundary between experiments and theory is our best bet for understanding computations and invariant representations in neural circuits.



L. Bernáez Timón, P. Ekelmans and N. Kravnyukova contributed equally to this work.

emergent network properties. With recent advances in experimental techniques for detailed anatomical reconstructions or large-scale activity recordings, rich biological data have become more available. The challenge when building network models is to reflect experimental results, either through a high level of detail or by finding an appropriate level of abstraction. Meanwhile, machine learning has facilitated the development of artificial neural networks, which are trained to perform specific tasks. While they have proven successful at achieving task-oriented behaviour, they are often abstract constructs that differ in many features from the physiology of brain circuits. Thus, it is unclear whether the mechanisms underlying computation in biological circuits can be investigated by analysing artificial networks that accomplish the same function but differ in their mechanisms. Here, we argue that building biologically realistic network models is crucial to establishing causal relationships between neurons, synapses, circuits and behaviour. More specifically, we advocate for network models that consider the connectivity structure and the recorded activity dynamics while evaluating task performance.

(Received 22 April 2022; accepted after revision 24 August 2022; first published online 7 September 2022)

Corresponding author T. Tchumatchenko: Institute of Experimental Epileptology and Cognition Research, University of Bonn Medical Centre, Bonn, Germany. Email: tatjana.tchumatchenko@uni-mainz.de

Abstract figure legend Neural networks execute the transformation of inputs into the neural activity that enables the execution of a task. Thus, they have the potential to establish causal relationships between the input, elements of a circuit, network activity and eventually behaviour. When the purpose of building a model is to understand the mechanisms that enable behaviour in biological circuits, the design of network models should be guided by three principles: (1) biological constraints are necessary (e.g. connectivity), (2) the network activity should match the recorded activity, and (3) the network should be able to perform the task of interest. Machine-learning-based training can help accomplish these goals.

Introduction

Seventy years ago, in 1952, Alan Hodgkin and Andrew Huxley concluded a series of papers about the flow of electric current through the membrane of a squid giant axon with the publication of a mathematical model describing the temporal shape of the action potential (Hodgkin & Huxley, 1952). At the time, this model constituted the first quantitative description of the ion channel activation underlying the action potential (or ‘spike’) generation in neurons. Yet, the Hodgkin and Huxley model was not the first model to describe a neuronal action potential in mathematical terms. A wealth of more abstract mathematical models that simulated the timing of neuronal spikes and the action potential spread across a network of neurons had been proposed long before Hodgkin and Huxley’s work. These included the work by Lapicque in 1907 in the form of an integrate-and-fire model (Abbott, 1999), McCulloch and Pitts in 1943, who modelled the activation of a neuron using a Boolean threshold function (McCulloch & Pitts, 1943), and Donald Hebb in 1949, who introduced the concept of synaptic plasticity as a regulator of spiking activity in his seminal book *The Organization of Behavior* (Hebb, 2005). These more abstract representations of neuronal activity equally became essential tools in studying complex networks in neuroscience.

Still today, the level of mathematical abstraction which is necessary and sufficient to understand the brain at the level of circuit computations and behaviour is a matter of debate (Gerstner & Naud, 2009; Pfeiffer & Pfeil, 2018). Overall, cognitive functions arise from the combination of multiple phenomena which occur at different spatial and temporal scales, ranging from ion channel dynamics to the activation of brain regions. Due to the complexity and interdependence of these phenomena, computational approaches in neuroscience often follow two complementary avenues: bottom-up and top-down. The ‘bottom-up’ approach starts from the observation of biophysical characteristics of neurons and synapses to build up mechanistically grounded models of the circuit activity following the premise that function will be an emergent property of the system. But what is the appropriate level of detail? For instance, how do the intricacies of ion channel dynamics put forward by Hodgkin and Huxley affect circuit-level phenomena? When can we abstract the ion channel kinetics and consider only the approximate time of a spike for all subsequent circuit-level analyses? Analogously, at the synaptic level, when is it crucial to detail the physiological processes governing short-term or long-term plasticity, and how can we capture the fundamental transformations in simpler mathematical

terms? Alternatively, the ‘top-down’ approach focuses on high-level cognitive functions or activity properties to reverse-engineer their underlying mechanisms and infer their physical implementation. Representative examples of this type of model are trained artificial neural networks (ANNs). Today, some of the trained ANNs that outperform humans on specific tasks (Silver et al. 2018; 2017) do not consider any detailed ion channel dynamics underlying a neural spike, nor do they learn according to the biologically reported local laws of synaptic plasticity. In general, these models bypass mechanistic descriptions of synaptic physiology (Abbott & Nelson, 2000; Tsodyks & Markram, 1997) and do not incorporate observed features of the dense recurrent cortical connectivity to reduce synaptic interactions to a set of machine learning optimization rules (Payeur et al. 2021; Rumelhart et al. 1986). At the neuronal level, these networks build only on concepts introduced before Hodgkin and Huxley’s 1952 model. Still, they are capable of learning complex tasks performed by different brain networks as well as producing neural dynamics that are similar to those recorded in experiments (Kar et al. 2019; Rajan et al. 2016; Sussillo et al. 2015). Thus, the question remains of when the detailed descriptions of neuronal and synaptic components can contribute novel aspects to our understanding of brain functions – both in bottom-up and top-down modelling efforts.

While acknowledging the power of more abstract models, we want to stress the importance of biological realism in understanding the mechanistic relationships between neurons, synapses, circuits and behaviour – in line with Hodgkin and Huxley’s core idea. A series of works comprising theory and experiments have illustrated impressively this idea by revealing, among others, the role of inhibition in stabilizing network activity (Tsodyks et al. 1997), the importance of recurrent connections in core object recognition (Kar & DiCarlo, 2021; Kar et al. 2019), or the synaptic mechanisms leading to the generation of sequential firing activity during decision-making (Rajan et al. 2016). We, therefore, believe that building biologically realistic neural circuit models that (1) capture experimentally reported circuit connectivity, (2) capture experimentally recorded neuronal and circuit activity, and (3) perform the behavioural task of interest is an important ongoing goal for computational neuroscience. But this comes at a price: the need for tractable theories that incorporate these elements and allow making testable hypotheses on the causal relationship between the building blocks of a neural circuit. In this review, we first provide an overview of the diversity of neural network models. We then illustrate the importance of linking the experimentally recorded connectivity and activity through network models. Finally, we review machine learning approaches and argue that training network models under realistic biological constraints can provide

insights into how biological neural networks can execute a function.

The essential building blocks of a neuronal network model

Network models represent abstractions of the physiology of biological neural circuits. They integrate the description of neurons and their connections through synapses in frameworks that allow studying the activity and function of neural circuits. Besides, network models enable us to establish causal relationships that are difficult to explore experimentally, infer mechanisms, dissect the role of specific circuit elements, or make experimentally testable hypotheses.

A neural network model of a brain circuit requires multiple building blocks (Fig. 1). Each of these elements can be modelled with varying degrees of abstraction. In general, the number of parameters and the model complexity increase as more biological details are included (Fig. 1, left to right). A complex model comes at the expense of mathematical tractability and computational efficiency. The modeller should therefore balance the benefit of a more biologically realistic model with the cost of a less tractable analysis. The choice of more abstract or detailed models depends on the scientific question to be answered, as different network features are known to be necessary for the emergence of specific properties. Thus, the modeller should decide which network properties are relevant for the study and which would obscure the mechanism of interest. This has been a challenge for generations of neuroscientists over the last decades (Almog & Korngreen, 2016; Pfeiffer & Pfeil, 2018) and has led to a wealth of models that compromise on different aspects of biological realism, recorded neural activity, or computational task performance.

A crucial element of a network model is the neuron (Fig. 1, Neuron). The most abstract model only distinguishes between two states, silent or active. Despite their simplicity, these so-called binary neurons can support the formation of network memory (Hopfield, 1982; McCulloch & Pitts, 1943). Neurons can also be modelled using continuous functions of the instantaneous state of their synaptic inputs. This is the case for the rectifier linear (ReLU) type of neuron (Hahnloser, 1998) or the sigmoidal neuron (Wilson & Cowan, 1972). These neuronal models are often used in ANNs and machine learning applications (Fukushima & Miyake, 1982; Minsky & Papert, 1969) as they suffice to support the network training in complex behavioural tasks (Mante et al. 2013). Next in complexity are integrate-and-fire models, where the neuronal activity depends on the recent synaptic input history and not only on its instantaneous state. The well-known

leaky integrate-and-fire (LIF) (Burkitt, 2006; Knight, 1972) and exponential integrate-and-fire (EIF) neurons (Fourcaud-Trocmé et al. 2003) belong to this type. In these models, the focus is on the spike time, so the shape of the action potential is not modelled. That is why LIF and EIF neurons are widely used in spiking network simulations that explore temporal features of the spiking activity, including the generation of network synchronous and asynchronous firing (Brunel, 2000; Renart et al. 2010). In contrast, when the scientific question requires capturing not only the spike time but also the temporal shape of the action potential, modelling ion channel kinetics and their conductance becomes necessary. Conductance-based neuronal models such as Hodgkin and Huxley's (1952) provide a description of ion channel dynamics and their interplay with the membrane potential. Network simulations of Hodgkin–Huxley neurons show that the action potential shape can have profound implications on the network dynamics, for instance, on de-synchronizing networks (Hesse et al. 2017). Often, these detailed conductance-based models require many parameters to describe ion channel physiology, which can mask the interpretation of the network simulation. To overcome this limitation, the FitzHugh–Nagumo model (FitzHugh, 1961) and the Izhikevich model (Izhikevich, 2003) provide a simplified phenomenological spiking implementation of the complex conductance-based models. Finally, the more detailed types of model consider the intricate neuronal morphologies by dividing the neuron into smaller inter-connected components (Edwards Jr & Mulloney, 1984; Hendrickson et al. 2011; Rall et al. 1992). These types of models have been successful at characterizing the spatial distribution of synaptic inputs and their consequences, which include the non-linear summation and temporal filtering of synaptic inputs at dendrites (Chavlis & Poirazi, 2021; Gidon et al. 2020; London & Häusser, 2005; Oberlaender et al. 2011).

In models, neurons that share common properties (i.e. the same type of synapses, similar receptive fields, among others) are often grouped into populations (Fig. 1, Populations). In the simplest case, all neurons are identical and make up a single population (Amit & Brunel, 1997; Brunel & Wang, 2003). However, single-population models of excitatory neurons are uncommon because it is impossible to stabilize the network activity to realistic, spontaneous firing rates (Amit & Brunel, 1997). Plus, brain circuits consistently contain both excitatory and inhibitory neurons, which cannot be assumed to be a single population since their activation leads to opposite effects (Douglas & Martin, 2004). Therefore, the study of neural network dynamics nowadays typically assumes two populations, one excitatory and one inhibitory (Brunel, 2000; Renart et al. 2010; van Vreeswijk & Sompolinsky, 1998; Wilson & Cowan, 1972). These excitatory–inhibitory network models can achieve physiological firing rates thanks to stabilization through recurrent inhibition (Amit & Brunel, 1997; Sadeh & Clopath, 2021; Tsodyks et al. 1997). Each of these excitatory and inhibitory populations can be broken down into subpopulations. One way to do so is by considering different cell types (Park & Geffen, 2020). Examples of these include networks with distinct inhibitory interneuron types (Del Molino et al. 2017; Litwin-Kumar et al. 2016; Mahrach et al. 2020). This type of model allows for dissecting the role of specific cell types in particular computations (Dipoppa et al. 2018; Keller et al. 2020; Litwin-Kumar et al. 2016). Furthermore, the inclusion of additional neural populations allows for a richer range of network dynamics (Hertäg & Sprekeler, 2019; Strogatz, 2018). Another way to increase the heterogeneity of a population is to imprint a distribution of receptive fields such as orientation selectivity (Chariker et al. 2016; Hennequin et al. 2018; Timón et al. 2022). In this case, defining a

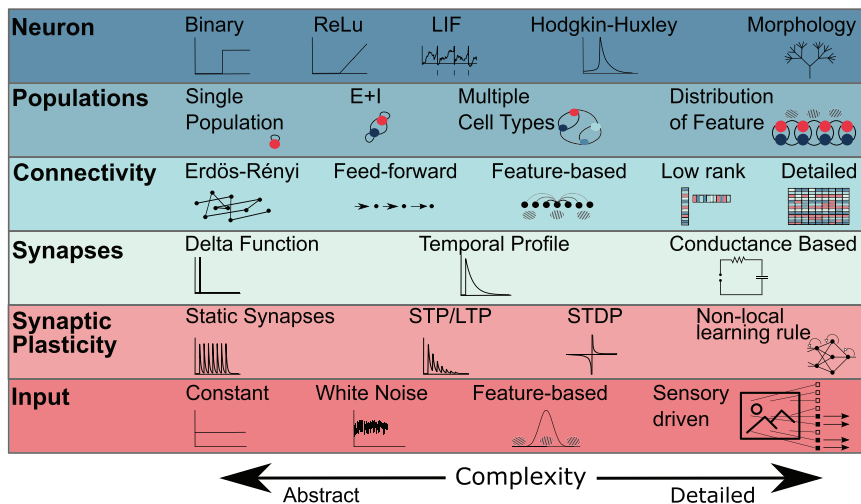


Figure 1. The essential building blocks of a neural network model ordered by degree of complexity

A network model representing the physiology of a neural circuit is characterized by its neuronal description, the categorization of neuronal properties into populations, the connectivity between neurons, the type of synapses and plasticity governing the synaptic strength, and the input that stimulates neurons. Each of these items can be modelled in different ways, with varying degrees of complexity.

single population with a continuum of neuronal features can suffice.

Another fundamental property of neural networks is their connectivity (Fig. 1, Connectivity). How neurons are connected within a network determines the network function and dynamics (Mastrogiuseppe & Ostojic, 2018; Ocker et al. 2017). In the simplest case, all neurons are connected to all. In such scenarios, all neurons share the same recurrent input, which promotes synchronicity (Brunel & Hansel, 2006; Wang & Buzsáki, 1996). More commonly used are the so-called randomly connected Erdős–Rényi networks (Brunel, 2000; Erdos & Rényi, 1959), in which pairs of neurons are connected with a certain fixed probability (Renart et al. 2010). These types of networks generate asynchronous states of activity, which can be described by low spiking correlations despite their neurons sharing considerable amounts of input (Renart et al. 2010). More realistic connectivity rules impose a higher connection probability for pairs of neurons that are closer topologically (Rosenbaum & Doiron, 2014) or in some feature space (Chariker et al. 2016; Timón et al. 2022), in line with experimental studies (Ko et al. 2011). This connectivity type leads to a strong correlation in the activity of similarly tuned neurons (Rosenbaum et al. 2017) and promotes the selective activation of clusters of neurons receptive to a given stimulus (Rosenbaum & Doiron, 2014). Structural connectivity can also be modelled using low-rank connectivity matrices, which are composed of a random plus a minimally structured part (Mastrogiuseppe & Ostojic, 2018). Studies have shown that such low-rank connectivity structures lead to low-dimensional dynamics (Mastrogiuseppe & Ostojic, 2018), in line with experimentally reported network activity (Mante et al. 2013). Finally, in cases in which the precise connectivity of the network is known from experiments, a detailed connectivity matrix can be incorporated in a model (Kunert et al. 2014; Varshney et al. 2011).

The next important characteristic of a neural network is its synapses. Synapses propagate the activity from spiking neurons to other neurons through complex chemical or electrical pathways. The result of this process is the depolarization or hyperpolarization of the postsynaptic neuron, depending on the type of presynaptic neuron. There are different ways to model the induced postsynaptic potential after a presynaptic spike (Fig. 1, Synapses). One important aspect of the synaptic current is its time course. In the simplest case, the membrane potential of the postsynaptic neuron is updated at a single time point, which can be modelled as a Dirac delta function (Brunel, 2000; Sanzeni Akitake et al. 2020; van Vreeswijk & Sompolinsky, 1998). On the other hand, the synaptic transmission can follow a more complex temporal profile where the synaptic input is distributed in time (Brunel & Wang, 2003; Renart et al. 2010). The

temporal properties of synaptic signals can affect the dynamical properties of the network activity, such as its rhythmicity (Brunel & Wang, 2003). Another essential property of the synapse is the effect that its activation induces on the postsynaptic neuron. Unlike current-based synaptic models, which treat synaptic transmission as a direct change in membrane potential to the postsynaptic neuron, a biophysical description of the ion channel physiology is included in conductance-based models (Cavallari et al. 2014). These models reproduce the change in conductance of the postsynaptic ion channels upon a presynaptic spike, which can lead to a non-linear summation of inputs since the current flow depends on the membrane potential of the neuron (Kuhn et al. 2004). The third fundamental property of the synapse is its capacity to change its connection strength. Dynamic or activity-dependent synaptic weights are used to model the ubiquitous phenomenon of synaptic plasticity (Fig. 1, Synaptic plasticity). In network models, synaptic plasticity leads to profound changes in the functional connectivity and allows networks to store memory (Mongillo et al. 2008) or adapt to perform certain tasks (Rajan et al. 2016). Classic examples of synaptic plasticity include calcium-mediated short-term plasticity (Tsodyks et al. 1998), spike-timing-dependent plasticity (Bi & Poo, 1998; Gerstner et al. 1996; Gjorgjieva et al. 2011; Kempter et al. 1999), or calcium and voltage-dependent synaptic plasticity (Bienenstock et al. 1982; Graupner & Brunel, 2012). It is worth noting that, even if the effect of these plasticity rules is local, i.e. they regulate the strength of a synapse based on the properties of the pre- and postsynaptic neurons, their impact manifests in the shape of different non-linearities at the network level (Mongillo et al. 2012; Timón et al. 2022; Zenke et al. 2015). Similarly, in the case of heterosynaptic plasticity, the strength of a synapse can be updated upon activation of other synapses of the same neuron. The inclusion of heterosynaptic plasticity rules has been linked to a homeostatic stabilization of the network activity (Chistiakova et al. 2015). A different class of plasticity includes the so-called non-local learning rules. These adapt the strength of synapses based on a global optimization target, which can be a behavioural output or a particular computational task (Kar et al. 2019; Mastrogiuseppe & Ostojic, 2018; Rajalingham et al. 2018; Sussillo et al. 2015). Examples of these include the backpropagation of error signals (Sacramento et al. 2018) or reward-driven learning (Foster et al. 2000; Frémaux et al. 2013; Seung, 2003; Soltani et al. 2006; Song et al. 2017). These global plasticity rules are common abstractions used in machine learning applications with great success at enabling networks to learn to execute a particular task. However, it is not clear how the information about the target or the global state of the circuit would be accessible to individual neurons and synapses in real biological circuits. Thus, translating

non-local into local synaptic learning rules that consider only the pre–post spike timing at a pair of neurons is now an active area of research showing promising findings (Payeur et al. 2021).

The last element of a neural network is the input (Fig. 1, Input). Besides the recurrent input from other neurons within the same circuit, neurons receive feed-forward input that originates from outside the modelled network. In biological circuits, the external input carries sensory-driven or otherwise processed information originating from other brain areas. The features of the modelled input should replicate the nature of such signals. The simplest type of external stimulation is a constant input, which resembles a sustained source of sensory stimulation and can be deterministic or stochastic (Brunel & Hansel, 2006; Wang & Buzs  ki, 1996). More complex input models are temporally dynamic to simulate a time-varying signal and are used to examine the dynamical properties of the network activity (Hahn et al. 2014; Roach et al. 2018). The external input can also be structured such that neurons with different receptive fields receive different inputs (Rosenbaum et al. 2017). Regardless of the specifics of the input, it is important to consider that its properties (i.e. statistics, temporal and spatial correlations) will be reflected in the activity of the modelled network. Indeed, the interplay between the recurrent and feed-forward inputs along the network hierarchy is key to understanding the generation of patterns of activity such as sequence generation (Long et al. 2010; Rajan et al. 2016), rhythmic activity (Brunel & Hakim, 1999; Brunel & Wang, 2003; Hesse et al. 2017), or network stability (Brunel, 2000; Mastrogiuseppe & Ostojic, 2018; van Vreeswijk & Sompolinsky, 1998).

All in all, the choice of the building blocks used in a network model requires a trade-off between a really detailed description of neural systems and a tractable model which can be easily manipulated to highlight its mechanisms (Fig. 1, complexity axis). In the limit of the most abstract network models, the system can be studied analytically without the need to simulate the network. These analytical descriptions are rate models that use a mean-field approximation to predict the average firing rate of neurons. This approach has been especially effective at identifying parameter regimes for the emergence of specific network states such as stable fixed points or chaos (van Vreeswijk & Sompolinsky, 1998), rhythmicity (Brunel, 2000), or various non-linearities (Ahmadian et al. 2013; Bienenstock et al. 1982; Kravnyukova & Tchumatchenko, 2018; Sanzeni Akitake et al. 2020). Regardless of the choice of model, the neural network operates a transformation from a given input into an output activity. This activity can be characterized by its mean firing rate, the precise timing of the emitted spikes, or the correlations between spike trains of different neurons. In the next section,

we discuss particular examples illustrating how network models can be used to infer the properties of a biological circuit from the recorded activity and how experimental connectomics data can constrain network model design.

Relating neural circuit activity and synaptic connectivity in models and experiments

Modern experimental techniques generate a steady stream of functional activity datasets, while progress in electron microscopy and automated image analysis provide previously inaccessible insights into the circuit connectome. These rapidly advancing connectivity datasets (Fig. 2A) can be related via neural network models of different complexity (Fig. 2B) to the recorded network activity (Fig. 2C). As a critical link between structure and function, network models can help make sense of complex experimental datasets and decipher the working principles of neural circuits. Here, we review achievements and challenges in modelling realistic neural networks aiming to bridge the growing number of activity and connectivity datasets.

During the last decades, neural activity recordings have made impressive progress. Calcium imaging techniques allow the monitoring of the activity of thousands of cells simultaneously (Pachitariu et al. 2017; Rose et al. 2014; Stringer et al. 2019) and even enable whole-brain activity recordings in freely moving animals (Nguyen et al. 2016) (Fig. 2C, middle). Similarly, high-density electrode arrays provide simultaneous recordings from hundreds of cells in deep brain areas (Jun et al. 2017; Steinmetz et al. 2021) (Fig. 2C, bottom). Both calcium imaging and electrophysiology techniques enable recordings of neural tuning curves in response to sensory stimuli or during behaviour, providing access to the functional role of neurons in the network (Fig. 2C, top). In addition, genetic markers can be used to classify the cell types in the recorded activity datasets (Fig. 2C, top). To reconstruct the circuit connectivity, multipatch-clamp recordings allow the estimation of the connection probability and synaptic strengths between specific cell types (Allen Institute for Brain Science, 2019; Cossell et al. 2015; Hofer et al. 2011; Jiang et al. 2015; Ko et al. 2011; Morgan et al. 2016; Seeman et al. 2018) (Fig. 2A, top). Calcium imaging *in vivo* preceding multipatch-clamp recordings or electron microscopy *in vitro* helps to identify neuronal response selectivity in the corresponding connectivity datasets (Bock et al. 2011; Cossell et al. 2015; Hofer et al. 2011; Ko et al. 2011; Microns Consortium, 2021) (Fig. 2A, middle). At the anatomical level, progress in electron microscopy and automated image analysis now provides detailed reconstructions of neuronal connections within a constrained volume of brain tissue (Helmstaedter et al. 2013; Oberlaender et al. 2012; Turner et al. 2022) (Fig. 2A, bottom), putting the first whole-brain

mammalian connectomes within reach (Abbott et al. 2020) (Fig. 2A, bottom). These newly acquired and growing experimental insights need to become a part of neural network models. At the same time, they call for previously developed network models to be reviewed.

Different studies have illustrated that neural network models can explain circuit mechanisms, provided their outcomes reproduce experimentally observed activity features. A prominent example demonstrating how activity recordings can be used to infer circuit properties concerns the study of the ‘paradoxical response’ using network models. The paradoxical response – a phenomenon observed across cortices – refers to the decrease in PV+ neurons’ activity after the stimulation of the PV+ population. The paradoxical response has been explained through inhibitory stabilization (Tsodyks et al. 1997), a regime which is well-understood in two population neural network models (Sadeh & Clopath,

2021; Sanzeni, Histed et al. 2020). Indeed, analyses of two population network models revealed connectivity regimes that relate the paradoxical response, the strength of recurrent connectivity between excitatory neurons, and inhibitory stabilization, providing a critical link between the circuit structure and function. Yet, since sensory cortices consist of multiple excitatory and inhibitory populations, the paradoxical response requires a better theoretical understanding of the rate and spiking network models with more than two neuronal populations (Fig. 2B). Recent theoretical studies on paradoxical effects in networks with multiple inhibitory subtypes in mouse V1 suggested that the stabilization of the circuit is primarily under the control of PV+ neurons. On the contrary, somatostatin and vasoactive intestinal peptide positive cells do not seem to impact the circuit’s stability significantly (Palmigiano et al. 2020; Sanzeni, Histed et al. 2020).

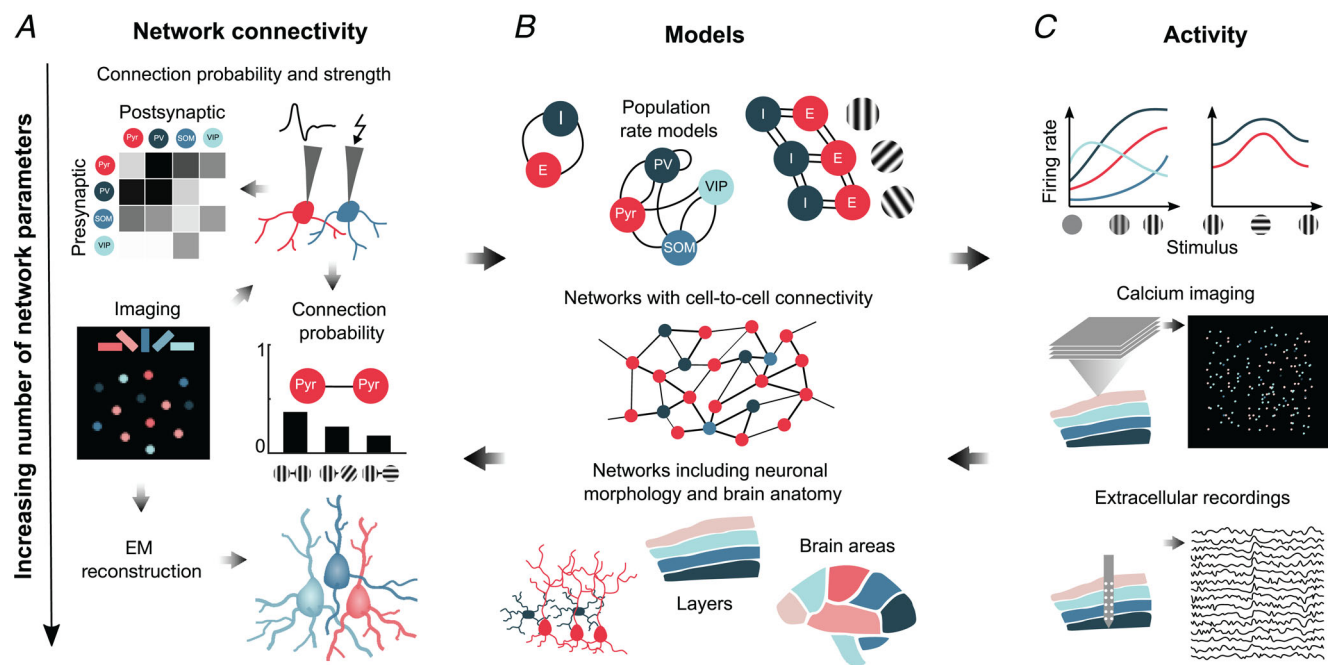


Figure 2. How to bridge network activity and connectivity using network models
 A, different levels of detail are available in the network connectivity reconstruction. Connection probability and strength (top, left) between neuronal populations enter the connectivity matrix of a neural network model. Neurons’ preferred orientation determined during *in vivo* calcium imaging recordings (middle, left) can be re-identified in subsequent multipatch-clamp recordings (middle, right) or automated electron microscopy (EM) image analyses (bottom), providing information on functional network connectivity. Connectomics methods now provide reconstructions of a cubic millimetre cortical volume containing tens of thousands of functionally identified neurons with hundreds of millions of synapses and include morphological details of individual cells in nanometre-scale resolution. B, neural network models can capture different levels of complexity in connectivity reconstructions and activity. Population rate models combine cells based on their genetic type or response properties (top). Spiking and rate networks containing individual neurons can capture heterogeneity in neuronal response properties or connectivity between individual neurons (middle). More detailed network models can reflect global network geometry and neuronal morphology. C, population activity (top) can be extracted from calcium imaging recordings of thousands of neurons on a small depth of the cortical surface (middle) or deep inter-layered recordings obtained from high-density multi-electrode arrays (bottom). A → C, inserting new detailed connectivity reconstructions into the network models requires their reassessment to reproduce the recorded activity. C → A, the network models can be used to infer the circuit connectivity from the recorded activity.

Beyond the paradoxical effect, computational studies have successfully shown how to infer neural circuit properties by exploiting the recorded activity dynamics, statistics and neuronal tuning curves. One elegant example of the latter is how the property of contrast invariance of orientation tuning (Anderson et al. 2000) constrained the architecture of network models. Imposing this property on the outcome of neuronal activity questioned the plausibility of a threshold linear transfer function previously used in V1 models. Subsequent work concluded that a neuronal power-law transfer function is the only function consistent with contrast invariance (Hansel & Van Vreeswijk, 2002; Miller & Troyer, 2002). The power-law supralinear relationship between the average neuronal membrane potential and its average spiking output was later confirmed in experiments (Priebe et al. 2004; Tan et al. 2011). The power-law transfer function has since become a part of the so-called stabilized supralinear network (SSN) models. The SSN models can reproduce numerous non-linear transformations observed in the visual cortex, such as surround suppression, variability quenching after stimulus onset, supersaturation and normalization (Ahmadian et al. 2013; Hennequin et al. 2018; Kraynyukova & Tchumatchenko, 2018; Rubin et al. 2015). A recent study (Renner et al. 2020) showed how to use the SSN model in combination with visual responses recorded in mouse V1 and thalamus to infer V1 connectivity weights between the populations of pyramidal and PV+ neurons. During the last decade, a series of connectivity measurements (Allen Institute for Brain Science, 2019; Hofer et al. 2011; Jiang et al. 2015; Ko et al. 2011) provided details on the connection probability and strength of synaptic connections in mouse V1. These studies revealed that the strength of the same type of connection could vary by an order of magnitude across experimental reports. The inference of connectivity via the SSN model showed that different connectivity configurations could approximate the same activity recordings. In other words, the SSN model helped in understanding that the connectivity configuration supporting the recorded activity is not necessarily unique, which is consistent with the variability in connectivity observed across prior experimental studies (Allen Institute for Brain Science, 2019; Hofer et al. 2011; Jiang et al. 2015; Ko et al. 2011). Despite the variability in experimentally reported connectivity, using the SSN model also revealed a systematic relationship among the magnitudes of the connectivity weights that appears necessary to generate the specific features of the recorded activity (Renner et al. 2020). Remarkably, this relationship between the connectivity weights was fulfilled in the individual connectivity studies. These results demonstrated how model-based connectivity inference could help in discovering ubiquitous connectivity motifs in otherwise diverse experimental measurements.

Given the available connectome datasets, it is tempting to assume that network models using bottom-up approaches, i.e. incorporating detailed neuronal and connectivity data Fig. 2A, will generate the recorded activity. But can a detailed network simulation based on state-of-the-art connectivity reconstructions directly reproduce all desired features of the recorded network activity? Over the last years, large-scale network simulations have been put forward (Arkhipov et al. 2018; Billeh et al. 2020; Markram et al. 2015) to relate detailed connectomics data to recorded activity. These simulations revealed that additional optimization of the connectivity weights is often required to generate biologically plausible activity which causes the weights to deviate from their initial biological values (Billeh et al. 2020). Here, detailed network models relating connectivity and recorded activity could help in understanding which connectivity features can vary without changing the relevant activity regime of the circuit and which are essential for behavioural outcomes and therefore remain invariant across experimental observations. In conclusion, neural network models are powerful tools that can relate the growing activity and connectivity datasets in an iterative process that involves critical and systematic revision of novel experimental insights. Past theoretical achievements demonstrate that continuously reviewing network models to make them consistent with updated experimental results is essential for understanding fundamental neural mechanisms.

Machine learning and the design of network models that can execute tasks

The ability to synthesize behaviour is a fundamental property of biological neural circuits that network models aim to capture. In the previous section, we have discussed how the exploration and understanding of fundamental neural mechanisms can be channelled via network models that link the connectivity datasets to the recorded activity (Fig. 2). But does recreating the recorded activity automatically serve task-oriented behaviour in network models? For example, would contrast-invariant activity automatically enable neural networks to perform contrast-invariant object recognition? From a computational perspective, it is still unclear how the accomplishment of a task follows from the activity and how to incorporate this fundamental property of brain circuits into neural network models. A promising approach to link connectivity and activity to behaviour in network models is to train the network parameters using machine learning. Powerful machine-learning methods are being increasingly used to train the connectivity of ANNs to perform behaviourally relevant tasks (Kar et al. 2019; Silver et al. 2017). This is because machine-learning

provides a set of implementable algorithms that reshape the network connectivity enabling the resulting network to execute specific task-related commands (Marblestone et al. 2016; Rumelhart & Zipser, 1985; Werbos, 1974). The objective of the network training depends on the scientific question and can focus on the optimization of the task performance, on the reproduction of a specific activity pattern, and many other factors (Fig. 3). Here, we review how trained ANNs can link connectivity, activity and behaviour from a computational perspective. Moreover, based on recent studies, we argue that including biological bottom-up constraints in the design of ANNs can synergistically reveal mechanisms underlying behaviour.

The machine-learning methodology has been applied to train abstract ANNs to execute complex tasks (Fig. 3A). Outstanding examples include the ANN AlphaZero, which outperforms humans in multiple two-player games (Silver et al. 2017, 2018), AlphaFold, which yields exceptional performance in the complex task of predicting protein folding (Eisenstein & others, 2021), or AlexNet, which marked a revolution in the field of computer vision (Krizhevsky et al. 2012). These types of neural networks

are extremely valuable from an engineering standpoint due to the high performance they achieve at executing the task they were designed for. At the same time, their connectivity, neuronal model or synapses differ greatly from those of biological neural networks, there is no guarantee that they achieve the task through mechanisms that resemble biology, and their activity may be difficult to interpret (Chakraborty et al. 2017). It is therefore unclear if abstract networks that are exclusively optimized for task performance can help understand behaviour in biological circuits. In particular, it is to be expected that they perform the task using functions that differ from biological transformations as they do not focus on recreating the activity dynamics recorded during behaviour.

ANNs trained as decoders can predict task outcomes from the recorded activity (Fig. 3B). This proves that there is information about the task encoded in the recorded activity and that this information is accessible using network models. For instance, recurrent ANNs trained to decode the arm position of a monkey from the recorded motor cortex activity (Sussillo et al. 2015) revealed that the arms position is encoded in a low-dimensional

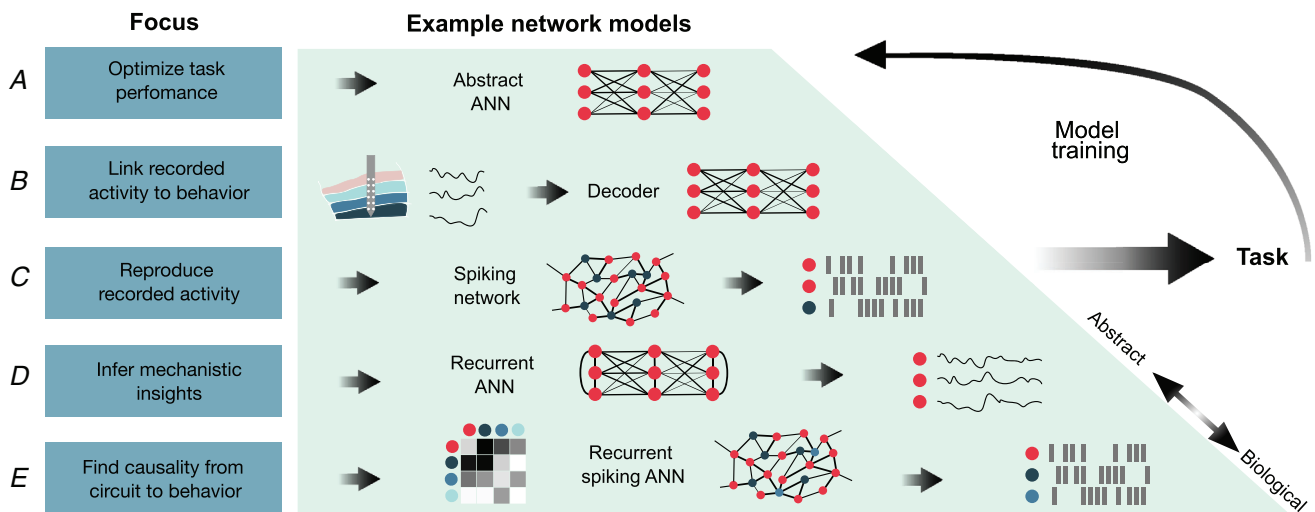


Figure 3. Diverse approaches to achieve task execution in bio-inspired network models using machine learning

Training the parameters of neural network models using machine learning can help to connect the input, the recorded activity, and the behavioural task of interest. The network architecture and the training can vary depending on the focus of the model and the level of biological detail (abstract ↔ biological axis). *A*, ANNs can be trained exclusively to optimize task performance. These models (e.g. AlphaZero) do not seek to understand the type of activity that enables the execution of the task and include few to no biological constraints. *B*, the activity recorded during behaviour can be fed into a neural network – decoder – to interpret behaviour. Decoders can predict behaviour based on the recorded activity; however, decoders do not address how neural circuits generate activity. *C*, the connectivity of spiking networks can be trained to generate a specific pattern of recorded activity. Once the network is trained, it is possible to study the mechanisms that underlie the generation of the activity in the model. But many models of this type do not address how the spiking activity relates to the task execution. *D*, the activity generated by ANNs trained in a certain task can be analysed retrospectively to look for mechanistic insights about the connectivity and activity underlying task performance. These insights can be used to guide specific experiments. *E*, biological constraints (e.g. detailed connectivity, spiking mechanism) known *a priori* can be included in the design of ANNs, which are then trained to execute a task. In this case, the network training on a task is subjected to the imposed biological constraints.

subspace of the motor cortex activity (Mante et al. 2013). Similarly, in the visual system, decoders could predict the behaviour of macaques performing an object categorization task from the recorded inferior temporal (IT) cortex population response (Majaj et al. 2015). This implies that the information contained in the activity of the IT cortex is sufficient to forecast object categorization behaviour in that task. In the context of navigation, it has been shown how the goal location can be decoded from the activity recorded in the orbitofrontal cortex of rats (Basu et al. 2021), which demonstrates that relevant information about the target location is encoded in the activity of that brain region. Hence, decoders can help elucidate from which brain structures task-related information can be recovered. At the same time, they provide insights into how the information is encoded, for instance, about its dimensionality. They do not, however, address the question of how the recorded activity is mechanistically generated.

The mechanisms underlying activity can be studied by dissecting artificial neural networks trained to recreate the activity recorded during task-related behaviour (Fig. 3C). Unlike the examples discussed before (Fig. 2), in which mathematically tractable models are used to reverse-engineer principles of the connectivity that underlies the recorded activity, here the potential connectivity supporting activity emerges as a result of the training. One example of training networks to study mechanisms applies to the generation of sequential firing observed in the posterior parietal cortex when an animal makes a left or right choice in the expectation of a reward (Harvey et al. 2012; Rajan et al. 2016). Training a randomly connected network to recreate the activity recorded during the task illustrated that only a small fraction of the recurrent connections need to undergo training to generate the measured sequential activation of neurons (Rajan et al. 2016). This means that sequential firing may be a feature of largely unstructured networks that emerges through learning. The learning algorithm had at least two constraints: the fraction of synapses that are trained can be controlled, and the change in the synaptic weight value is limited. The key message here is that training networks to recreate recorded activity and subjecting the training to certain biological constraints can reveal causal relationships between the connectivity and the activity. However, it is not clear what the functional implications are of the generated activity in the context of task execution.

One of the first attempts to build models that simultaneously address how the activity is generated from the connectivity and how it is related to the execution of a task (Fig. 3D) comes from the field of vision. These models focused on training the network connectivity to optimize core object recognition. Object recognition is the ability to categorize objects and

to identify them at different positions, in different sizes, light conditions or contexts (Grill-Spector et al. 2001). An early representative example of this class of models is a hierarchical feed-forward ANN, the so-called 'Neocognitron' (Fukushima & Miyake, 1982). The novelty of this model relied on its layered structure that allows subsequent extraction of features of the input (Grill-Spector et al. 2001). The Neocognitron acquired the ability to perform core object recognition through unsupervised learning. Interestingly, an emergent property of the Neocognitron is that the structure of its activity patterns shares similarities with that of the visual system, for instance with regard to the encoding of different features of the stimulus across cortical layers (DiCarlo & Cox, 2007; Felleman & Van Essen, 1991; Fukushima & Miyake, 1982; Kobatake & Tanaka, 1994). This suggests that the separation of feature encoding across layers may be an optimal network organization in the context of visual perception. Since the Neocognitron, more complex network models have been developed which can recognize objects and at the same time generate responses that match the recorded activity of real neurons to newly presented stimuli (Horikawa & Kamitani, 2017; Yamins et al. 2014). These trained models have provided mechanistic insights into the relationship between activity and behaviour and inspired interesting experiments (Fig. 3D). For instance, training a hierarchical network model on object recognition allowed inferring from the model the pixel patterns that would activate neurons in a particular fashion across the model layers (Bashivan et al. 2019). This model-based prediction was validated in monkeys, which demonstrated that groups of neurons can be selectively activated or silenced using specifically designed visual stimuli (Bashivan et al. 2019). There are further examples of how these network models can inspire experiments (Kar & DiCarlo, 2021; Kar et al. 2019). In these studies (Kar & DiCarlo, 2021; Kar et al. 2019), the authors compared the performance of monkeys and trained ANNs in classifying images and identified types of images for which ANNs perform significantly worse than monkeys (Kar et al. 2019). After measuring the activity of neurons in the monkey brain in response to those challenging images, they detected a latency in the neuronal response compared to the representation of the images that were not challenging to the model (Kar et al. 2019). A compelling hypothesis was that recurrent feedback, which was not a feature of the feed-forward ANNs, was the cause of the latency. To test this hypothesis, parts of the macaque ventrolateral prefrontal cortex, which is a recurrent node, were inactivated. This experiment revealed that macaques without recurrency show deficits in object recognition tasks (Kar & DiCarlo, 2021). From a computational perspective, it was shown that including recurrent connections in the ANN model improved the model performance at recognizing the images that

were challenging for a purely feed-forward network (Kietzmann et al. 2019). Furthermore, recurrent models produce neuronal responses that are more similar to IT cortex than the previous feed-forward ANNs (Kar et al. 2019; Liang & Hu, 2015). Overall, these results indicate that recurrent connections are critical for efficient object recognition. Moreover, they provide evidence that a closed-loop dialogue between experiments and ANN modelling can generate experimentally testable hypotheses and highlight the biological features that improve the performance of neural networks.

Building upon the importance of including biological principles, recent work has aimed at integrating detailed biological constraints into the design of ANNs, merging bottom-up and top-down modelling (Fig. 3E). One example is the use of experimentally reported connectomes of primates to constrain the connectivity and the initial connection strengths of ANNs trained to perform a working-memory task (Goulas et al. 2021). Compared to a randomly structured ANN, these connectivity constraints did not compromise task performance, which is significant as the inclusion of biological constraints reduces the available parameter space for training. Others have studied the functional effects of the network topology by imposing biologically realistic connectivities and found that modularity is key at driving performance in a memory-encoding task (Suárez et al. 2021). Besides the connectivity, trained ANNs can incorporate biological insights in the form of neuronal signalling. Unlike classical ANNs, which consist of static neurons with a time-invariant activation function, neurons in so-called spiking neural networks transfer information in the form of discrete time-resolved spikes. Theoretical work has developed effective training algorithms which take into account the spike-timing of individual neurons (Tavanaei et al. 2019). This is a promising approach because it remains unclear whether information in the brain is encoded through firing rates or spike times (Brette, 2015). Concerning the training, it has been recently noted that the widespread use of gradient descent to train ANNs violates Dale's law as it can change the sign of the weights connecting two neurons throughout the training process (Cornford et al. 2021). This would imply that the same neuron can be either excitatory or inhibitory as a function of the training state, which is not consistent with experimental observations. New studies have proposed novel training algorithms to preserve these principles without sacrificing task performance (Cornford et al. 2021). Indeed, the inclusion of biological constraints such as the spiking model, the connectivity and the learning rule into the bottom-up design of ANNs is essential to ensure that, after training, the network performs the function through mechanisms that resemble biology. Thus, combining bottom-up modelling with the top-down training that

enables ANNs to execute tasks is a promising approach to elucidate the mechanism underlying function in biological neural networks.

Conclusion

The brain is a complex biological system that processes information through billions of neurons organized in intricate networks. Due to this complexity, theoretical models of neural networks are often used in order to elucidate how it operates. Over the last decades, many types of models have been developed to explore or mimic the mechanisms that underpin the brain's computational capabilities, such as decision making, navigation, visual perception and many more. This multitude of models ranges from abstract and mathematically tractable equations to highly complex multidimensional dynamical systems that take into account detailed biological processes on different temporal and spatial scales (Fig. 1). Neural network models can shed light on multiple levels of the mechanisms underlying brain functions, such as relating the structure of a network to its activity (Fig. 2) or providing insights into the algorithmic implementation of complex behavioural tasks (Fig. 3). Some of these models are built bottom-up from the observation of biophysical features at the neuronal and synaptic levels to design networks and study their emergent properties. Others reverse-engineer possible network implementations permitting known activity features or cognitive functions in a top-down fashion. Irrespective of how abstract or realistic these models are, they can all advance our knowledge of brain function through their ability to capture the key transformations in mathematical terms. For instance, models can unveil what are the minimal network requirements for the emergence of certain properties (Hopfield, 1982; McCulloch & Pitts, 1943) or can provide connectivity configurations that support the execution of a task (Russakovsky et al. 2015; Silver et al., 2017, 2018). Moreover, evidence shows that identifying the shortcomings of neural network models can be instrumental in revealing essential mechanisms of brain function (Kar & DiCarlo, 2021; Kar et al. 2019). This illustrates how a careful analysis of the limitations of a model, even when it appears to underperform, can lead to experimentally testable hypotheses and provide insights into the principles underlying activity and behaviour.

Multiple studies have shown that biologically realistic network models can be crucial to uncovering core principles of brain function by mediating an effective cross-talk between experiments, bottom-up and top-down design (Bashivan et al. 2019; Kar et al. 2019; Kietzmann et al. 2019; Liang & Hu, 2015; Rajan et al. 2016). We have reviewed how these principles can guide the design of neural network models (Figs 2 and 3). In

order to promote such cross-talk between these different frameworks, research can follow different strategies. One strategy is to impose strict biological constraints during the network training process (Fig. 3E). This strategy relies on the development of training algorithms that account for spiking activity (Tavanaei et al. 2019), that conform to Dale's law (Cornford et al. 2021) or that employ biologically inspired synaptic plasticity rules. An alternative strategy is to use unconstrained network training to highlight the network configurations which can enable a behaviour of interest. Empirical constraints and bottom-up modelling can then help discriminate which implementations are biologically feasible *a posteriori*. Regardless of the strategy used, a combined framework bridging biophysical mechanisms and cognitive functions will likely face the problem of interpretability. Even if all parameters of a high dimensional ANN are known, its huge number often hinders the understanding of how these networks perform their tasks. Another challenge to overcome in studying biological circuits using machine learning algorithms is that biological networks are subjected to multiple competing considerations which ANNs typically do not optimize for. For example, the optimal implementation of a biological circuit for a given task would presumably compromise task performance with other objectives such as efficient energy consumption, wiring length, processing speed, the ability to learn from limited data, or robustness to noise and errors (Pallasdies et al. 2021). In this context, using ANNs to predict the optimal implementation of biological networks would require defining and quantifying these competing factors. The endeavour of bridging the bottom-up and top-down approaches of neural network modelling is confronted with many challenges. Yet, far from preventing scientific progress, these challenges seem to be helping define the right questions and encouraging computational neuroscientists to develop novel solutions.

To sum up, we argue that to gain a mechanistic understanding of brain function and of emergent circuit phenomena it is essential to equip neural network models with biological connectivity as well as to evaluate if the modelled activity matches recorded activity. Training these biologically constrained networks to perform tasks can help elucidate the mechanisms underlying behaviour in neural circuits. This cross-talk can now be fuelled by recent advances in connectomics, electrophysiology and machine-learning.

References

- Abbott, L. F. (1999). Lapicque's introduction of the integrate-and-fire model neuron (1907). *Brain Research Bulletin*, **50**(5–6), 303–304.
- Abbott, L. F., Bock, D. D., Callaway, E. M., Denk, W., Dulac, C., Fairhall, A. L., Fiete, I., Harris, K. M., Helmstaedter, M., Jain, V., Kasthuri, N., LeCun, Y., Lichtman, J. W., Littlewood, P. B., Luo, L., Maunsell, J. H. R., Reid, R. C., Rosen, B. R., Rubin, G. M., ..., & Van Essen, D. C. (2020b). The mind of a mouse. *Cell*, **182**(6), 1372–1376.
- Abbott, L. F., & Nelson, S. B. (2000). Synaptic plasticity: Taming the beast. *Nature Neuroscience*, **3**(S11), 1178–1183.
- Ahmadian, Y., Rubin, D. B., & Miller, K. D. (2013). Analysis of the stabilized supralinear network. *Neural Computation*, **25**(8), 1994–2037.
- Allen Institute for Brain Science (2019). Synaptic physiology coarse matrix dataset Available from: <https://brainmap.org/explore/connectivity/synaptic-physiology>
- Almog, M., & Korngreen, A. (2016). Is realistic neuronal modeling realistic? *Journal of Neurophysiology*, **116**(5), 2180–2209.
- Amit, D. J., & Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, **7**, 237–252.
- Anderson, J. S., Lampl, I., Gillespie, D. C., & Ferster, D. (2000). The contribution of noise to orientation tuning in cat visual cortex. *Science*, **290**(5498), 1968–1972.
- Arkipov, A., Gouwens, N. W., Billeh, Y. N., Gratiy, S., Iyer, R., Wei, Z., Xu, Z., Abbasi-Asl, R., Berg, J., Buice, M., Cain, N., da Costa, N., de Vries, S., Denman, D., Durand, S., Feng, D., Jarsky, T., Lecoq, J., Lee, B., ..., & Koch, C. (2018). Visual physiology of the layer 4 cortical circuit in silico. *PLoS Computational Biology*, **14**(11), e1006535.
- Bashivan, P., Kar, K., & DiCarlo, J. J. (2019). Neural population control via deep image synthesis. *Science*, **364**(6439), eaav9436.
- Basu, R., Gebauer, R., Herfurth, T., Kolb, S., Golipour, Z., Tchumatchenko, T., & Ito, H. T. (2021). The orbitofrontal cortex maps future navigational goals. *Nature*, **599**(7885), 449–452.
- Bi, G. Q., & Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, **18**(24), 10464–10472.
- Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, **2**(1), 32–48.
- Billeh, Y. N., Cai, B., Gratiy, S. L., Dai, K., Iyer, R., Gouwens, N. W., Abbasi-Asl, R., Jia, X., Siegle, J. H., Olsen, S. R., Koch, C., Mihalas, S., & Arkipov, A. (2020). Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. *Neuron*, **106**(3), 388–403.e18.
- Bock, D. D., Lee, W. C. A., Kerlin, A. M., Andermann, M. L., Hood, G., Wetzel, A. W., Yurgenson, S., Soucy, E. R., Kim, H. S., & Reid, R. C. (2011). Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, **471**(7337), 177–182.
- Brette, R. (2015). Philosophy of the spike: Rate-based vs. spike-based theories of the brain. *Frontiers in Systems Neuroscience*, **9**, 151.

- Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience*, **8**(3), 183–208.
- Brunel, N., & Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Computation*, **11**(7), 1621–1671.
- Brunel, N., & Hansel, D. (2006). How noise affects the synchronization properties of recurrent networks of inhibitory neurons. *Neural Computation*, **18**(5), 1066–1110.
- Brunel, N., & Wang, X. J. (2003). What determines the frequency of fast network oscillations with irregular neural discharges? i. synaptic dynamics and excitation-inhibition balance. *Journal of Neurophysiology*, **90**(1), 415–430.
- Burkitt, A. N. (2006). A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biological Cybernetics*, **95**(1), 1–19.
- Cavallari, S., Panzeri, S., & Mazzoni, A. (2014). Comparison of the dynamics of neural interactions between current-based and conductance-based integrate-and-fire recurrent networks. *Frontiers in Neural Circuits*, **8**, 12.
- Chakraborty, S., Tomsett, R., Raghavendra, R., Harborne, D., Alzantot, M., Cerutti, F., Srivastava, M., Preece, A., Julier, S., Rao, R. M., Kelley, T. D., Braines, D., Sensoy, M., Willis, C. J., & Gurrum, P. (2017). Interpretability of deep learning models: A survey of results. *IEEE* pp. 1–6.
- Chariker, L., Shapley, R., & Young, L. S. (2016). Orientation selectivity from very sparse lgn inputs in a comprehensive model of macaque v1 cortex. *Journal of Neuroscience*, **36**(49), 12368–12384.
- Chavlis, S., & Poirazi, P. (2021). Drawing inspiration from biological dendrites to empower artificial neural networks. *Current Opinion in Neurobiology*, **70**, 1–10.
- Chistiakova, M., Bannon, N. M., Chen, J. Y., Bazhenov, M., & Volgushev, M. (2015). Homeostatic role of heterosynaptic plasticity: Models and experiments. *Frontiers in Computational Neuroscience*, **9**, 89.
- Cornford, J., Kalajdzievski, D., Leite, M., Lamarquette, A., Kullmann, D. M., & Richards, B. (2021). Learning to live with dale's principle: Anns with separate excitatory and inhibitory units. *bioRxiv*. <https://doi.org/10.1101/2020.11.02.364968>
- Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., Hofer, S. B., & Mrsic-Flogel, T. D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*, **518**(7539), 399–403.
- Del Molino, L. C. G., Yang, G. R., Mejias, J. F., & Wang, X. J. (2017). Paradoxical response reversal of top-down modulation in cortical circuits with three interneuron types. *eLife*, **6**, e29742.
- DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, **11**(8), 333–341.
- Dipoppa, M., Ranson, A., Krumin, M., Pachitariu, M., Carandini, M., & Harris, K. D. (2018). Vision and locomotion shape the interactions between neuron types in mouse visual cortex. *Neuron*, **98**(3), 602–615.e8.
- Douglas, R. J., & Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annual Review of Neuroscience*, **27**(1), 419–451.
- Edwards, Jr D. H., & Mulloney, B. (1984). Compartmental models of electrotonic structure and synaptic integration in an identified neurone. *The Journal of Physiology*, **348**(1), 89–113.
- Eisenstein, M. et al. (2021). Artificial intelligence powers protein-folding predictions. *Nature*, **599**(7886), 706–708.
- Erdos, P., & Rényi, A. (1959). On random graphs I. *Publicationes Mathematicae Debrecen*.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, **1**, 1–47.
- FitzHugh, R. (1961). Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, **1**(6), 445–466.
- Foster, D. J., Morris, R. G., & Dayan, P. (2000). A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus*, **10**(1), 1–16.
- Fourcaud-Trocmé, N., Hansel, D., Van Vreeswijk, C., & Brunel, N. (2003). How spike generation mechanisms determine the neuronal response to fluctuating inputs. *Journal of Neuroscience*, **23**(37), 11628–11640.
- Frémaux, N., Sprekeler, H., & Gerstner, W. (2013). Reinforcement learning using a continuous time actor-critic framework with spiking neurons. *PLoS Computational Biology*, **9**(4), e1003024.
- Fukushima, K., & Miyake, S. (1982). Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In: Amari, S., Arbib, M.A. (eds) *Competition and Cooperation in Neural Nets. Lecture Notes in Biomathematics*, vol **45**. Springer, Berlin, Heidelberg, 267–285. https://doi.org/10.1007/978-3-642-46466-9_18
- Gerstner, W., Kempter, R., Van Hemmen, J. L., & Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature*, **383**(6595), 76–78.
- Gerstner, W., & Naud, R. (2009). How good are neuron models? *Science*, **326**(5951), 379–380.
- Gidon, A., Zolnik, T. A., Fidzinski, P., Bolduan, F., Papoutsis, A., Poirazi, P., Holtkamp, M., Vida, I., & Larkum, M. E. (2020). Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science*, **367**(6473), 83–87.
- Gjorgjieva, J., Clopath, C., Audet, J., & Pfister, J. P. (2011). 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, USA*, **108**(48), 19383–19388.
- Goulas, A., Damicelli, F., & Hilgetag, C. C. (2021). Bio-instantiated recurrent neural networks: Integrating neurobiology-based network topology in artificial networks. *Neural Networks*, **142**, 608–618.
- Graupner, M., & Brunel, N. (2012). Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *Proceedings of the National Academy of Sciences, USA*, **109**(10), 3991–3996.
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, **41**(10–11), 1409–1422.

- Hahn, G., Bujan, A. F., Frégnac, Y., Aertsen, A., & Kumar, A. (2014). Communication through resonance in spiking neuronal networks. *PLoS Computational Biology*, **10**(8), e1003811.
- Hahnloser, R. L. (1998). On the piecewise analysis of networks of linear threshold neurons. *Neural Networks*, **11**(4), 691–697.
- Hansel, D., & Van Vreeswijk, C. (2002). How noise contributes to contrast invariance of orientation tuning in cat visual cortex. *Journal of Neuroscience*, **22**(12), 5118–5128.
- Harvey, C. D., Coen, P., & Tank, D. W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, **484**(7392), 62–68.
- Hebb, D. O. (2005). *The organization of behavior: A neuro-psychological theory*. Psychology Press.
- Helmstaedter, M., Briggman, K. L., Turaga, S. C., Jain, V., Seung, H. S., & Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, **500**(7461), 168–174.
- Hendrickson, E. B., Edgerton, J. R., & Jaeger, D. (2011). The capabilities and limitations of conductance-based compartmental neuron models with reduced branched or unbranched morphologies and active dendrites. *Journal of Computational Neuroscience*, **30**(2), 301–321.
- Hennequin, G., Ahmadian, Y., Rubin, D. B., Lengyel, M., & Miller, K. D. (2018). The dynamical regime of sensory cortex: Stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron*, **98**(4), 846–860.e5.
- Hertäg, L., & Sprekeler, H. (2019). Amplifying the redistribution of somato-dendritic inhibition by the interplay of three interneuron types. *PLoS Computational Biology*, **15**(5), e1006999.
- Hesse, J., Schleimer, J. H., & Schreiber, S. (2017). Qualitative changes in phase-response curve and synchronization at the saddle-node-loop bifurcation. *Physical Review E*, **95**(5), 052203.
- Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, **117**(4), 500–544.
- Hofer, S. B., Ko, H., Pichler, B., Vogelstein, J., Ros, H., Zeng, H., Lein, E., Lesica, N. A., & Mrsic-Flogel, T. D. (2011). Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature Neuroscience*, **14**(8), 1045–1052.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA*, **79**(8), 2554–2558.
- Horikawa, T., & Kamitani, Y. (2017). Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications*, **8**(1), 1–15.
- Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, **14**(6), 1569–1572.
- Jiang, X., Shen, S., Cadwell, C. R., Berens, P., Sinz, F., Ecker, A. S., Patel, S., & Tolias, A. S. (2015). Principles of connectivity among morphologically defined cell types in adult neocortex. *Science*, **350**(6264), aac9462.
- Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydın, Ç., Barbic, M., Blanche, T. J., Bonin, V., Couto, J., Dutta, B., Gratiy, S. L., Gutnisky, D. A., Häusser, M., Karsh, B., ..., & Harris, T. D. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, **551**(7679), 232–236.
- Kar, K., & DiCarlo, J. J. (2021). Fast recurrent processing via ventrolateral prefrontal cortex is needed by the primate ventral stream for robust core visual object recognition. *Neuron*, **109**(1), 164–176. e5.
- Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., & DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience*, **22**(6), 974–983.
- Keller, A. J., Dipoppa, M., Roth, M. M., Caudill, M. S., Ingrosso, A., Miller, K. D., & Scanziani, M. (2020). A disinhibitory circuit for contextual modulation in primary visual cortex. *Neuron*, **108**(6), 1181–1193. e8.
- Kempler, R., Gerstner, W., & Van Hemmen, J. L. (1999). Hebbian learning and spiking neurons. *Physical Review E*, **59**(4), 4498.
- Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences, USA*, **116**(43), 21854–21863.
- Knight, B. W. (1972). Dynamics of encoding in a population of neurons. *Journal of General Physiology*, **59**(6), 734–766.
- Ko, H., Hofer, S. B., Pichler, B., Buchanan, K. A., Sjöström, P. J., & Mrsic-Flogel, T. D. (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature*, **473**(7345), 87–91.
- Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, **71**(3), 856–867.
- Krainyukova, N., & Tchumatchenko, T. (2018). Stabilized supralinear network can give rise to bistable, oscillatory, and persistent activity. *Proceedings of the National Academy of Sciences, USA*, **115**(13), 3464–3469.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, **25**, 1097–1105.
- Kuhn, A., Aertsen, A., & Rotter, S. (2004). Neuronal integration of synaptic input in the fluctuation-driven regime. *Journal of Neuroscience*, **24**(10), 2345–2356.
- Kunert, J., Shlizerman, E., & Kutz, J. N. (2014). Low-dimensional functionality of complex network dynamics: Neurosensory integration in the caenorhabditis elegans connectome. *Physical Review E*, **89**(5), 052805.
- Liang, M., & Hu, X. (2015). Recurrent convolutional neural network for object recognition In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3367–3375.
- Litwin-Kumar, A., Rosenbaum, R., & Doiron, B. (2016). Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes. *Journal of Neurophysiology*, **115**(3), 1399–1409.

- London, M., & Häusser, M. (2005). Dendritic computation. *Annual Review of Neuroscience*, **28**(1), 503–532.
- Long, M. A., Jin, D. Z., & Fee, M. S. (2010). Support for a synaptic chain model of neuronal sequence generation. *Nature*, **468**(7322), 394–399.
- Mahrach, A., Chen, G., Li, N., van Vreeswijk, C., & Hansel, D. (2020). Mechanisms underlying the response of mouse cortical networks to optogenetic manipulation. *eLife*, **9**, e49967.
- Majaj, N. J., Hong, H., Solomon, E. A., & DiCarlo, J. J. (2015). Simple learned weighted sums of inferior temporal neuronal firing rates accurately predict human core object recognition performance. *Journal of Neuroscience*, **35**(39), 13402–13418.
- Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, **503**(7474), 78–84.
- Marblestone, A. H., Wayne, G., & Kording, K. P. (2016). Toward an integration of deep learning and neuroscience. *Frontiers in Computational Neuroscience*, **10**, 94.
- Markram, H., Müller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., Kahou, G. A. A., Berger, T. K., Bilgili, A., Buncic, N., Chalimourda, A., Chindemi, G., Courcol, G. D., Delalondre, F., Delattre, V., ..., & Schürmann, F. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell*, **163**(2), 456–492.
- Mastrogiuseppe, F., & Ostojic, S. (2018). Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, **99**(3), 609–623. e29.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, **5**(4), 115–133.
- Microns Consortium (2021). Microns explorer. <https://www.microns-explorer.org/>
- Miller, K. D., & Troyer, T. W. (2002). Neural noise can explain expansive, power-law nonlinearities in neural response functions. *Journal of Neurophysiology*, **87**(2), 653–659.
- Minsky, M., & Papert, S. (1969). Perceptron: an introduction to computational geometry. MIT Press
- Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, **319**(5869), 1543–1546.
- Mongillo, G., Hansel, D., & Van Vreeswijk, C. (2012). Bistability and spatiotemporal irregularity in neuronal networks with nonlinear synaptic transmission. *Physical Review Letters*, **108**(15), 158101.
- Morgan, J. L., Berger, D. R., Wetzel, A. W., & Lichtman, J. W. (2016). The fuzzy logic of network connectivity in mouse visual thalamus. *Cell*, **165**(1), 192–206.
- Nguyen, J. P., Shipley, F. B., Linder, A. N., Plummer, G. S., Liu, M., Setru, S. U., Shaevitz, J. W., & Leifer, A. M. (2016). Whole-brain calcium imaging with cellular resolution in freely behaving *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences, USA*, **113**(8), E1074–E1081.
- Oberlaender, M., Boudewijns, Z. S., Kleele, T., Mansvelter, H. D., Sakmann, B., & de Kock, C. P. (2011). Three-dimensional axon morphologies of individual layer 5 neurons indicate cell type-specific intracortical pathways for whisker motion and touch. *Proceedings of the National Academy of Sciences, USA*, **108**(10), 4188–4193.
- Oberlaender, M., de Kock, C. P., Bruno, R. M., Ramirez, A., Meyer, H. S., Dercksen, V. J., Helmstaedter, M., & Sakmann, B. (2012). Cell type-specific three-dimensional structure of thalamocortical circuits in a column of rat vibrissa cortex. *Cerebral Cortex*, **22**(10), 2375–2391.
- Ocker, G. K., Hu, Y., Buice, M. A., Doiron, B., Josic, K., Rosenbaum, R., & Shea-Brown, E. (2017). From the statistics of connectivity to the statistics of spike times in neuronal networks. *Current Opinion in Neurobiology*, **46**, 109–119.
- Pachitariu, M., Stringer, C., Dipoppa, M., Schröder, S., Rossi, L. F., Dalgleish, H., Carandini, M., & Harris, K. D. (2017). Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *BioRxiv*. <https://doi.org/10.1101/061507>
- Pallasdies, F., Norton, P., Schleimer, J. H., & Schreiber, S. (2021). Neural optimization: Understanding trade-offs with pareto theory. *Current Opinion in Neurobiology*, **71**, 84–91.
- Palmigiano, A., Fumarola, F., Mossing, D. P., Kravnyukova, N., Adesnik, H., & Miller, K. D. (2020). Structure and variability of optogenetic responses identify the operating regime of cortex. *bioRxiv*. <https://doi.org/10.1101/2020.11.11.378729>
- Park, Y., & Geffen, M. N. (2020). A circuit model of auditory cortex. *PLoS Computational Biology*, **16**(7), e1008016.
- Payeur, A., Guerguiev, J., Zenke, F., Richards, B. A., & Naud, R. (2021). Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits. *Nature Neuroscience*, **24**(7), 1010–1019.
- Pfeiffer, M., & Pfeil, T. (2018). Deep learning with spiking neurons: Opportunities and challenges. *Frontiers in Neuroscience*, **12**, 74.
- Priebe, N. J., Mechler, F., Carandini, M., & Ferster, D. (2004). The contribution of spike threshold to the dichotomy of cortical simple and complex cells. *Nature Neuroscience*, **7**(10), 1113–1122.
- Rajalingham, R., Issa, E. B., Bashivan, P., Kar, K., Schmidt, K., & DiCarlo, J. J. (2018). Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *Journal of Neuroscience*, **38**(33), 7255–7269.
- Rajan, K., Harvey, C. D., & Tank, D. W. (2016). Recurrent network models of sequence generation and memory. *Neuron*, **90**(1), 128–142.
- Rall, W., Burke, R. E., Holmes, W. R., Jack, J. J., Redman, S. J., & Segev, I. (1992). Matching dendritic neuron models to experimental data. *Physiological Reviews*, **72**(4), S159–S186.
- Renart, A., De La Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., & Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science*, **327**(5965), 587–590.
- Renner, S., Kravnyukova, N., Bauer, Y., Born, G., Kotkat, A. H., Liu, X., Spacek, M., Tushev, G., Tchumatchenko, T., & Busse, L. (2020). Inference of network connectivity from responses to briefly flashed gratings in mouse V1 using a stabilized supralinear network model (SSN). *Bernstein Conference Abstract*, **16**, 10–11.
- Roach, J. P., Pidde, A., Katz, E., Wu, J., Ognjanovski, N., Aton, S. J., & Zochowski, M. R. (2018). Resonance with sub-threshold oscillatory drive organizes activity and optimizes learning in neural networks. *Proceedings of the National Academy of Sciences, USA*, **115**(13), E3017–E3025.

- Rose, T., Goltstein, P. M., Portugues, R., & Griesbeck, O. (2014). Putting a finishing touch on GECIs. *Frontiers in Molecular Neuroscience*, **7**, 88.
- Rosenbaum, R., & Doiron, B. (2014). Balanced networks of spiking neurons with spatially dependent recurrent connections. *Physical Review X*, **4**(2), 021039.
- Rosenbaum, R., Smith, M. A., Kohn, A., Rubin, J. E., & Doiron, B. (2017). The spatial structure of correlated neuronal variability. *Nature Neuroscience*, **20**(1), 107–114.
- Rubin, D. B., Van Hooser, S. D., & Miller, K. D. (2015). The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, **85**(2), 402–417.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, **323**(6088), 533–536.
- Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science*, **9**(1), 75–112.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, **115**(3), 211–252.
- Sacramento, J., Ponte Costa, R., Bengio, Y., & Senn, W. (2018). Dendritic cortical microcircuits approximate the back-propagation algorithm. *Advances in Neural Information Processing Systems*, **31**, 8735–8746.
- Sadeh, S., & Clopath, C. (2021). Inhibitory stabilization and cortical computation. *Nature Reviews. Neuroscience*, **22**(1), 21–37.
- Sanzeni, A., Akitake, B., Goldbach, H. C., Leedy, C. E., Brunel, N., & Histed, M. H. (2020). Inhibition stabilization is a widespread property of cortical networks. *eLife*, **9**, e54875.
- Sanzeni, A., Histed, M. H., & Brunel, N. (2020). Response nonlinearities in networks of spiking neurons. *PLoS Computational Biology*, **16**(9), e1008165.
- Seeman, S. C., Campagnola, L., Davoudian, P. A., Hoggarth, A., Hage, T. A., Bosma-Moody, A., Baker, C. A., Lee, J. H., Mihalas, S., Teeter, C., Ko, A. L., Ojemann, J. G., Gwinn, R. P., Silbergeld, D. L., Cobbs, C., Phillips, J., Lein, E., Murphy, G., Koch, C., Zeng, H., & Jarsky, T. (2018). Sparse recurrent excitatory connectivity in the microcircuit of the adult mouse and human cortex. *eLife*, **7**, e37349.
- Seung, H. S. (2003). Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. *Neuron*, **40**(6), 1063–1073.
- Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, **362**(6419), 1140–1144.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of go without human knowledge. *Nature*, **550**(7676), 354–359.
- Soltani, A., Lee, D., & Wang, X. J. (2006). Neural mechanism for stochastic behaviour during a competitive game. *Neural Networks*, **19**(8), 1075–1090.
- Song, H. F., Yang, G. R., & Wang, X. J. (2017). Reward-based training of recurrent neural networks for cognitive and value-based tasks. *eLife*, **6**, e21492.
- Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., Chen, S., Colonell, J., Gardner, R. J., Karsh, B., Kloosterman, F., Kostadinov, D., Mora-Lopez, C., O'Callaghan, J., Park, J., ..., & Harris, T. D. (2021). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science*, **372**(6539), eabf4588.
- Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., & Harris, K. D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, **364**(6437), eaav7893.
- Strogatz, S. H. (2018). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC press.
- Suárez, L. E., Richards, B. A., Lajoie, G., & Misić, B. (2021). Learning function from structure in neuromorphic networks. *Nature Machine Intelligence*, **3**(9), 771–786.
- Sussillo, D., Churchland, M. M., Kaufman, M. T., & Shenoy, K. V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. *Nature Neuroscience*, **18**(7), 1025–1033.
- Tan, A. Y., Brown, B. D., Scholl, B., Mohanty, D., & Priebe, N. J. (2011). Orientation selectivity of synaptic input to neurons in mouse and cat primary visual cortex. *Journal of Neuroscience*, **31**(34), 12339–12350.
- Tavanaei, A., Ghodrati, M., Kheradpisheh, S. R., Masquelier, T., & Maida, A. (2019). Deep learning in spiking neural networks. *Neural Networks*, **111**, 47–63.
- Timón, L. B., Ekelmans, P., Konrad, S., Nold, A., & Tchumatchenko, T. (2022). Synaptic plasticity controls the emergence of population-wide invariant representations in balanced network models. *Physical Review Research*, **4**(1), 013162.
- Tsodyks, M., Pawelzik, K., & Markram, H. (1998). Neural networks with dynamic synapses. *Neural Computation*, **10**(4), 821–835.
- Tsodyks, M. V., & Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the National Academy of Sciences, USA*, **94**(2), 719–723.
- Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J., & McNaughton, B. L. (1997). Paradoxical effects of external modulation of inhibitory interneurons. *Journal of Neuroscience*, **17**(11), 4382–4388.
- Turner, N. L., Macrina, T., Bae, J. A., Yang, R., Wilson, A. M., Schneider-Mizell, C., Lee, K., Lu, R., Wu, J., Bodor, A. L., Bleckert, A. A., Brittain, D., Froudarakis, E., Dorkenwald, S., Collman, F., Kemnitz, N., Ih, D., Silversmith, W. M., Zung, J., ..., & Seung, H. S. (2022). Reconstruction of neocortex: Organelles, compartments, cells, circuits, and activity. *Cell*, **185**(6), 1082–1100.e24.
- van Vreeswijk, C., & Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Computation*, **10**(6), 1321–1371.

- Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H., & Chklovskii, D. B. (2011). Structural properties of the caenorhabditis elegans neuronal network. *PLoS Computational Biology*, *7*(2), e1001066.
- Wang, X. J., & Buzsáki, G. (1996). Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model. *Journal of Neuroscience*, *16*(20), 6402–6413.
- Werbos, P. (1974). *Beyond regression: "new tools for prediction and analysis in the behavioral sciences"*. Ph. D. dissertation, Harvard University.
- Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*(1), 1–24.
- Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences, USA*, *111*(23), 8619–8624.
- Zenke, F., Agnes, E. J., & Gerstner, W. (2015). Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communications*, *6*(1), 1–13.

Additional information

Competing interests

The authors declare no competing interests.

Author contributions

L.B.T., P.E., N.K. and T.T. generated the first draft of the paper with input from all authors: abstract and introduction (P.E., L.B.T., T.T.), section on the essential building blocks (P.E., T.T., L.B.T.), section on neural circuit activity and connectivity in models and experiments (N.K., T.R., T.T.), section on machine learning and the design of network models (L.B.T., L.B., T.R., P.E.) and conclusion (L.B.T., P.E., T.T.). All authors provided critical feedback, suggested literature, and helped shape the

content. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All persons designated as authors qualify for authorship and all those who qualify for authorship are listed.

Funding

This work was supported by the German Research Foundation (DFG) SPP2041 (T.T., L.B.), DFG BU 1808/6-1 and BU 1808/6-2 (L.B.) and TC 67/4-1 and TC 67/4-2 (T.T.), University of Bonn Medical Centre, University of Mainz Medical Centre, Loewe Centre for Multiscale Modelling in Life Sciences (T.T.) and DFG SFB 1233 (TP10, TP13; project number: 276 693 517; L.B.), SFB 1080 (T.T.) and SFB 1089 (T.T. and T.R.) and by an Add-on fellowship of the Joachim Herz Stiftung (L.B.T.).

Acknowledgements

T.T., T.R. and L.B.T. thank all our group members for fruitful discussions and comments on the manuscript.

Open access funding enabled and organized by Projekt DEAL.

Keywords

ANN, bio-inspired, machine learning, modelling, neural network, neuroscience, task execution, training

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the HTML view of the article. Supporting information files available:

Peer Review History