Opinion

Cell PRESS

Challenges of understanding brain function by selective modulation of neuronal subpopulations

Arvind Kumar, Ioannis Vlachos, Ad Aertsen, and Clemens Boucsein

Bernstein Center Freiburg (BCF), and Neurobiology and Biophysics, Faculty of Biology, University of Freiburg, Freiburg, Germany

Neuronal networks confront researchers with an overwhelming complexity of interactions between their elements. A common approach to understanding neuronal processing is to reduce complexity by defining subunits and infer their functional role by selectively modulating them. However, this seemingly straightforward approach may lead to confusing results if the network exhibits parallel pathways leading to recurrent connectivity. We demonstrate limits of the selective modulation approach and argue that, even though highly successful in some instances, the approach fails in networks with complex connectivity. We argue to refine experimental techniques by carefully considering the structural features of the neuronal networks involved. Such methods could dramatically increase the effectiveness of selective modulation and may lead to a mechanistic understanding of principles underlying brain function.

Introduction

Structural features of networks on multiple scales

The mammalian brain is often referred to as the most complex biological system, not only because of the large number of cells ($\sim 10^{11}$) but mainly because of the multitude of interactions between them. Even if all connections ($\sim 10^{15}$) were known, a system with such complexity could not be successfully treated at all levels of detail simultaneously. Instead, it is advisable to focus on an appropriate spatial scale and to reduce the lower levels to putative functional units, approximating their intrinsic fine structure with the help of fewer variables, if not ignoring them entirely.

At the macroscopic scale ($\geq$1 cm), the brain can be described as a network of different areas or regions. Together with this anatomical modularity, neurophysiological evidence indicates that these different regions in the brain are to some extent specialized to process specific sensory, motor, or cognitive information, and that they are interconnected in a non-trivial fashion [1,2]. At mesoscopic scales ($\sim$1 cm), brain regions are themselves often anatomically segregated and can be subdivided into layers, subfields, or nuclei. Examples are the hippocampal formation, the interconnected cortical layers or columns, and the networks in the limbic system such as the basal ganglia and amygdala. Thus, at macro and mesoscopic levels the brain is best considered as a network of networks (NoN). Finally, at microscopic scales ($\sim$1 mm), brain tissue can be defined as a network of neurons, which can be broadly categorized as excitatory and inhibitory. At this spatial resolution, brain tissue conforms most appropriately to the term 'network' because the interconnected nodes are naturally defined (individual neurons), and their connectivity can, to some extent, be measured experimentally. Below this scale, specialized, spatially extended cells can add additional degrees of freedom through compartmentalization and non-linear dendritic integration [3,4] which might affect computations performed at higher levels. However, although different spatial scales of network organization render the system complex, the presence of specific network motifs (Figure 1) raises the hope that different spatial scales can be described using common principles.

Such conceptual reduction of complexity is commonly referred to as model-building. Their mathematical formulation makes them accessible for theoretical studies, opening the opportunity to study their behavior beyond the limitations that often constrain experiments on real brains. Commonly used models typically fall into one of three classes: feedforward models, ring models, and models with recurrent connectivity (Figure 1). The simplest model is the feedforward network, which allows an intuitive understanding of its working principles merely from visual inspection of its graphical description (Figure 2A). In such a feedforward network it is straightforward to characterize the separate contributions of all populations by controlled modulation of their activity.

Several networks in the sensory and motor periphery may conform (at least approximately) to such simple architecture regarding their intra- and inter-network connectivity – hence they can be treated as effectively feedforward. Examples include cortico-striatal interactions as well as many networks in the sensory and motor periphery, and the hippocampus with its largely feedforward, trisynaptic pathway. In fact, at short time-scales (such that the neuronal activity cannot travel over the full loop), even the entorhinal cortex-hippocampal and the thalamo-cortical loop (thalamus and layer IV) can be studied as an effectively feedforward network. Finally, particular random recurrent networks can also be reduced to an effective feedforward structure [5].

Corresponding authors: Kumar, A. (arvind.kumar@biologie.uni-freiburg.de); Boucsein, C. (clemens.boucsein@biologie.uni-freiburg.de).

0166-2236/$ – see front matter
 http://dx.doi.org/10.1016/j.tins.2013.06.005

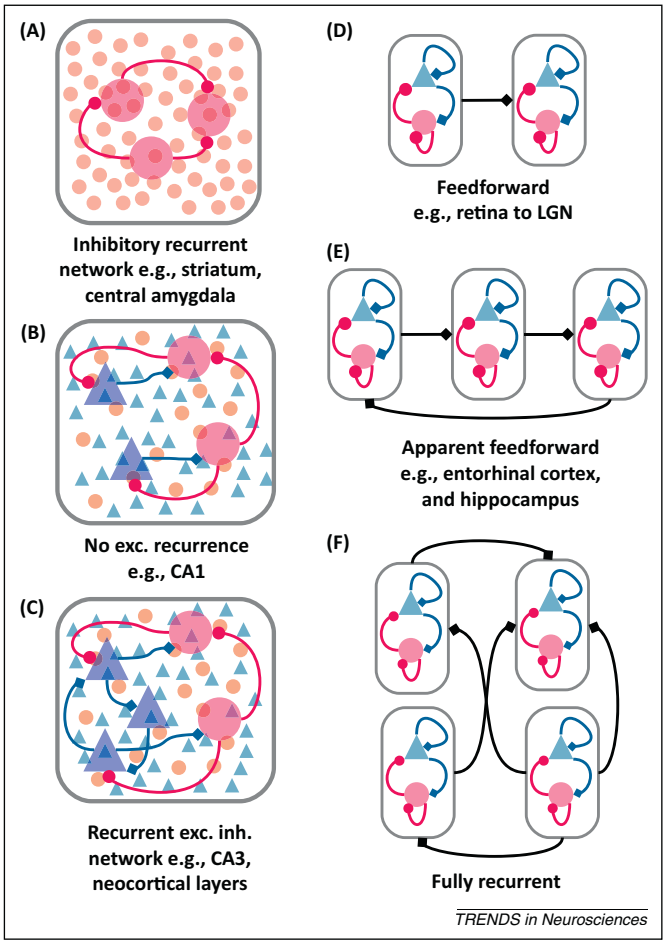

Figure 1. Organization of neuronal networks at different spatial scales. **(A–C)** Neuronal network at microscopic scales (~1 mm). There are three main motifs at this scale: (A) a recurrent network composed solely of inhibitory neurons; for example, the striatum or the central amygdala. The neurons are maintained in their spiking state by excitation from other networks. (B) A recurrent network composed of both excitatory and inhibitory neurons but with little or no recurrent connectivity among excitatory neurons; for example, the CA1 subfield of the hippocampus. (C) A recurrent network with both excitatory and inhibitory neurons and mutual connectivity between the two populations; for example, the individual layers of the neocortex and the CA3 subfield of the hippocampus. **(D–F)** Beyond the scale of 1 mm, networks in the brain are in fact networks of networks (NoN). These NoNs can be classified into three different categories as follows. (D) Feedforward connection between two networks is the simplest NoN motif. Such motifs are most common in the sensory and motor periphery. (E) Next, NoNs such as the entorhinal and hippocampus loop are effectively feedforward, although there may be multiple projections from one field to other fields (e.g., entorhinal cortex connects to dentate gyrus, CA3, and CA1). (F) Finally, in fully recurrent NoNs the constituent neuronal networks (nodes) form both feedforward and feedback connections. Such network motifs are commonly observed both at anatomical (e.g., the inter-layer connectivity in the neocortex) and functional levels (e.g., the network of various subnetworks involved in visual information processing). Abbreviations: exc., excitatory; exc. inh., excitatory and inhibitory; LGN, lateral geniculate nucleus.

With recent improvements in brain stimulation technology it has become possible to modulate neuronal networks and/or groups of neurons with unprecedented specificity. Because selective stimulation of neuronal networks has been very informative regarding the functional interactions in early sensory and peripheral motor circuits (essentially feedforward structures), it seems only natural to use these tools in the same approach to other brain areas as well, and to all levels of structural detail, even down to the single cell level. However, several considerations concerning the dynamics of network activity reveal fundamental problems in extending this seemingly successful approach to networks that show a slightly more complex connectivity structure beyond simple feedforward circuitry.

Limitations of the selective activity modulation approach

Beyond the first stages of sensory processing or the penultimate stages of motor processing, most networks in the brain cannot be approximated by a feedforward structure. Higher brain areas exhibit more recurrency for which it is non-trivial to reveal the specific activity patterns that implement a presumed function and to identify the elements involved. To demonstrate the problems that may occur in recurrent networks, we converted our feedforward network into a recurrent network by adding one more connection from the third to the second population (Figure 2A,B). In this model it is necessary to modulate the activity of P2 and P3 both separately and simultaneously to understand their functional roles in the network.

In the case of more interacting neuronal populations or neurons (both referred to as nodes in the sequel), modulating the activity of single nodes within the network yields only limited information about network function [6,7] – when activity in one node inhibits a second and excites a third, and these in turn are coupled and feed back onto the first, even brief consideration of the network behavior reveals its complexity. Thus, for a network of five interacting nodes, to assess the functional contributions of its elements and their possible interactions it would be necessary to modulate the activity of one, two, three, and four nodes simultaneously, and to record the corresponding network activity in each case. In general, for a network of N nodes, we would need to selectively modulate all single nodes, and all combinations of two to $(N-1)$ nodes, separately to identify the relevant subset(s) of nodes (cf. [8]). For a network of N interacting elements, the number of all possible subsets of nodes is given by Bell's number [9]:

$$B_N = \sum_{k=0}^{N-1} \frac{(N-1)!}{k!(N-1-k)!} B_k \qquad [1]$$

Unfortunately, B_N grows faster than exponentially with N. For a system consisting of six populations (the minimal population size to study thalamo-cortical interactions), $B_6 = 203$, and for a system with ten populations, $B_{10} = 115\ 975$. Thus, as the number of interacting nodes in a network grows, the number of node combinations that need to be modulated rapidly becomes prohibitively large.

For example, even considering only five layers of a neocortical column (those containing the majority of all cell bodies), to comprehend the contribution of each layer to a particular function we would need $B_5 = 52$ different combinations of single and multi-layer stimulations. Obviously, to treat neocortical networks that way is an oversimplification, and ignores that a large fraction of the inputs originate from the surround [10]. Moreover, interlayer connectivity [11] implies that activity dynamics and layer-specific responses are influenced by the activity of various neuron types and networks in different layers [12,13]. Thus, modulation of a specific neuron type in a

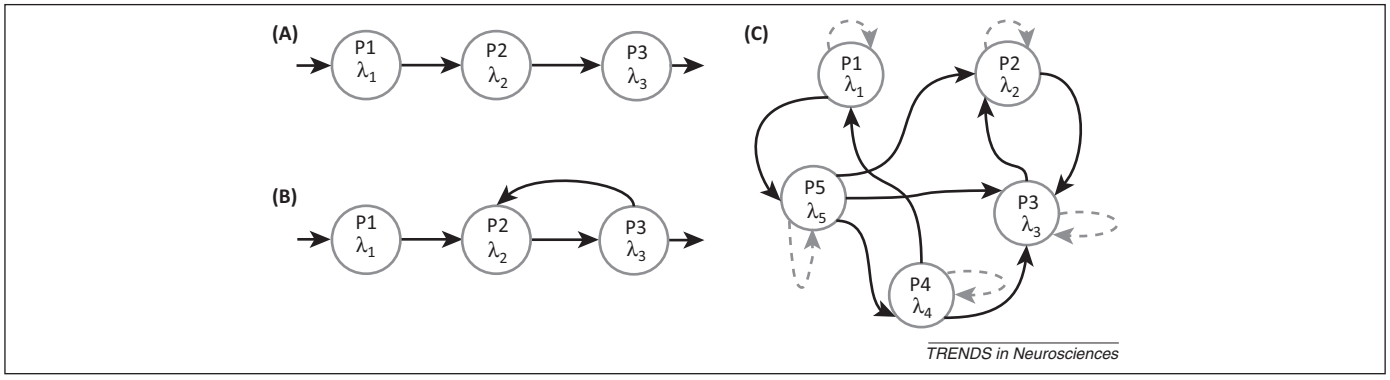

Figure 2. Only simple networks of interacting neuronal populations can be studied using selective modulation of individual populations. **(A)** In a feedforward network it is straightforward to characterize and understand the contribution of each population onto the next. **(B)** Any deviation from the strict feedforward motif requires control of the activity of more than one population. In this three-population network we need to stimulate, in addition to single populations, pairs of populations to understand network function. **(C)** In a recurrent network of networks (NoN) (Figure 1D) with *N* populations, we need B_N (Equation 1) different combinations of simultaneous stimulation of multiple populations to understand the impact of one population upon the others. Unfortunately, the number B_N grows with *N* faster than exponentially. In the toy example with five populations, this requires 52 different stimulation patterns involving 1–5 populations. Abbreviation: λ_{1-5}, firing rates of respective neuron populations P1–P5.

given layer [14] is not sufficient to understand the dynamics [12] and stimulus response [13] of a cortical column. Indirectly, this issue was already encountered when Lee *et al.* [15] attempted to extract the contribution of specific pyramidal neurons to the fMRI–BOLD (blood oxygen level-dependent) signal, a problem complicated even more by recurrent inter- and intra-layer connections [16].

This combinatorial problem becomes even more intractable when nodes interact in a non-linear fashion (as is most often the case in brain networks) or when couplings are dynamic and activity-dependent. In both cases, one would need to linearize the system and perform the experiments and associated analyses around a specific operating point – for each of which Bell's number applies – and, hence, to understand the full system, Bell's number would need to be multiplied by the number of operating points chosen.

Finally, we note that most experiments thus far rely on steady-state responses upon stimulation of one or more nodes. However, when dealing with interactions among dynamical systems it is important to consider both transient and steady-state responses to identify the functional subset of the network because steady-state responses alone can be misleading [17].

Taken together, these considerations demonstrate that, even if a thorough understanding of the functional network mechanisms is not mandatory (e.g., in medical applications), the sheer number of possibilities to affect network behavior by selective modulation of its elements renders a successful outcome of such experiments rather unlikely.

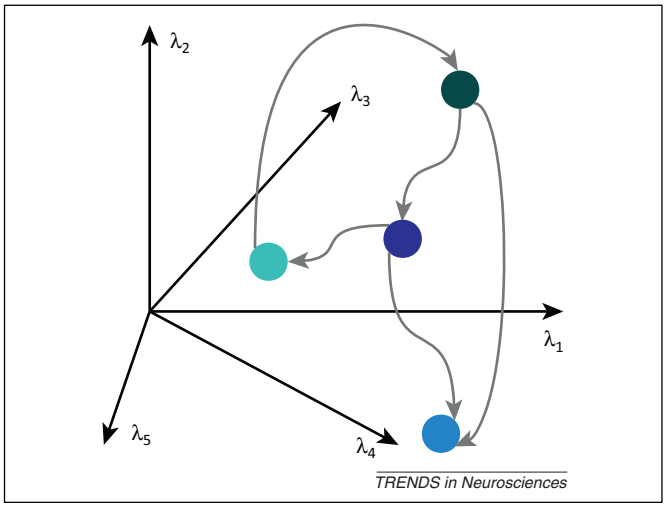

Figure 3. Scheme of the dynamical space of a five-population network of networks (NoN). The population activity represents the average activity of neurons in the given neuronal population. In a dynamical system description, the function of the NoN, in an abstract sense, can be described as the transition from one state to another (i.e., control of the network activity dynamics). Here the states are schematically indicated by colored circles; arrows mark the transitions between states. Indeed, often more than one transition from any given state is possible. Controllability provides the necessary criteria to determine whether a system is controllable; in other words, whether it can be 'steered' from one dynamical state to another in finite time. Abbreviation: λ_{1-5}, firing rates of different neuron populations.

Common techniques for selective activity modulation

Assuming that the above-mentioned combinatorial issues with selective modulation could be resolved (see below for new perspectives), another important issue is the availability of experimental methods for selectively modulating the activity of neuronal networks. Deciding which methods to apply for studying a given network partly depends on the research goals: if we wish to understand thoroughly the working principles underlying both, the function and the functioning of the network, we need to identify the functional role of each element and characterize their interactions. However, for the technical task of controlling of prosthetic devices by brain activity, local field potentials (LFPs) can be utilized [18], even though the mechanisms underlying their generation are still under debate [19] and the LFP is unlikely to affect brain activity (but how LFP might influence spiking see [20,21]).

The earliest successful selective modulation experiments probably date back to early Greco-Roman times when Galen (AD 129–199) discovered that mental impairments of gladiators were associated with specific head injuries. Since those early, involuntary brain lesion studies, experimenters have sought to refine loss-of-function approaches and, indeed, lesions of brain structures by targeted ablations at different levels of detail have provided important insights into the functional organization of

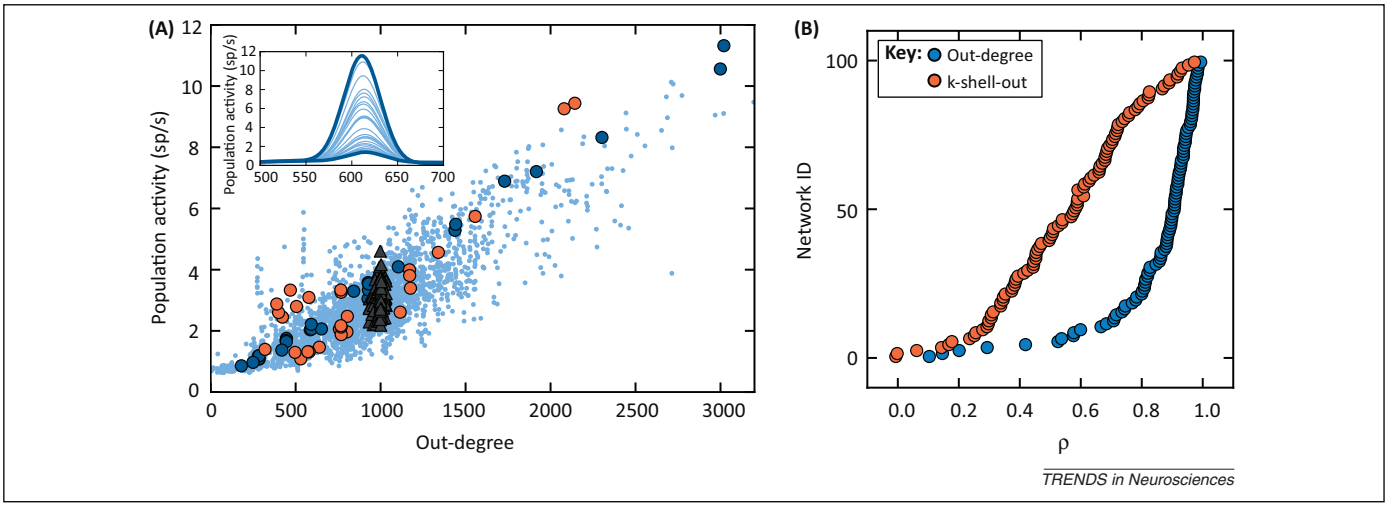

Figure 4. Estimation of embeddedness in numerical simulations of spiking neuronal networks. (A, inset) Network response (peri-stimulus time histogram, PSTH) for identical stimulation of 30 different subpopulations of 250 neurons each in an example network. (A) We estimated the sum of the PSTHs for 100 different networks (see [42] for details) upon stimulation of different subpopulations (250 neurons). Two networks with small-world properties are highlighted (dark blue, orange dots). The random networks which do not show much out-degree variance are indicated by triangles. The network response increases with the out-degree, but the same out-degree can also give different responses. (B) Average correlation coefficient (ρ, sorted) between the sum of PSTHs as a function of out-degree and k-shell-out indices. Both these measures predict the network response; however, neither alone describes it completely. (Figure adapted from [42]). Abbreviation: sp/s, spikes per second.

the CNS [22]. First, pieces of brain tissue were physically removed and, later, silenced by injection of toxic chemicals or local tissue cooling (Table 1). A major refinement in selective modulation came with electrical stimulation. Since 1870 [23] it is used both to identify the function of brain areas and as a therapy to intervene with aberrant activity dynamics associated with brain disorders. Moreover, electrical stimulation also allows activation rather than silencing of nerve tissue. However, the effects of electrical brain stimulation (including behavioral effects) are generally not well understood, partly because they may involve a multitude of pathways [24].

All these 'classical' approaches affect the activity of more or less the entire cell population in a locally confined volume of brain tissue, most likely (through axonal stimulation) together with cells downstream and upstream. This implies that they operate on multiple levels simultaneously, involving a hierarchy of networks. With recent developments in optogenetic methods these caveats have been largely eliminated, leading to an unprecedented richness of tools for selective activation/inactivation. Transgenic or viral transfection with light-gated ion channels now allow a specific group of neurons to be stimulated while recording neuronal activity without stimulus artifacts [25,26]. Thus, it seems only natural to combine the formerly successful modulation approaches with these new capabilities. In the light of the aforementioned theoretical considerations, however, we should appreciate that the simple modulation approach will rarely (only in structures reducible to feedforward networks) lead to reliable guidelines for clinical intervention, let alone to a thorough understanding of the neuronal mechanisms involved.

What is a network, and how can its function be investigated?

Defining networks

Even assuming we could deal with the combinatorial explosion associated with selective modulation, another fundamental issue needs to be resolved: how to define the network which should be stimulated or whose function needs to be understood? Classical methods used to modulate brain activity activate/inactivate cells in a spatially localized fashion. Hence, it seems intuitive to consider anatomically proximate neurons as a functional unit. Indeed, neighboring neurons often appear to be responsive to similar stimulus features, or are co-active during similar tasks – as expressed in the concept of 'functional maps'. However, experimental data show that neurons involved in cognitive and motor functions are distributed throughout the brain [27]. Likewise, primary sensory areas can be modulated by other stimulus modalities [28–31], and spatial grouping of neurons with similar stimulus preferences does not seem to be a general feature of cortical networks [32]. The applicability of proximity-based grouping breaks down entirely if cell-specific markers are employed, as in many optogenetic approaches. Here, only a subpopulation of cells within a given volume is modulated based on the expression of particular biomarkers. At the level of such subpopulations, feedforward circuits have been identified [33], but the respective cells are also involved in other circuits within the same tissue volume and, thus, modulating their activity may have confounding effects.

In addition, it can be problematic to demarcate functionally specialized neuronal subpopulations from the outset. Various classification schemes based on neuron location, morphology, gene expression and spike patterns have been suggested [34–37]. Undoubtedly, many of these features will relate to the functional role of the respective populations: activating excitatory cells has very different effects on the network than activating inhibitory neurons. Also, applying different criteria may lead to overlapping boundaries between putative neuron types, and differentiation is still limited for key neuronal populations. For instance, the subdivision of neocortical pyramidal cells based on their firing pattern seems to be of limited

Table 1. Selective activity modulation approaches

Stimulation method	Temporal resolution	Spatial resolution	Specificity
Electrical microstimulation	Good	Poor, because axons from distal regions can also be stimulated	Poor. To some extent it is possible to select between stimulating somata and axons [62]
Chemical injection, e.g., neurotransmitter agonists and antagonists	Poor	Depends on the diffusion of the chemical	Good
Magnetic stimulation (e.g., transcranial magnetic stimulation, TMS)	Good	Poor, because axons from distal regions can also be stimulated	Poor
Cryogenic (cooling of brain tissue)	Good	Poor, because axons from distal regions can also be affected	Poor
Optogenetic stimulation	Good. However, thus far all transfected neurons are stimulated simultaneously [63]	Moderate. Depends on the penetration of light and expression pattern of the optogenetic tools. Potential for improvement by two-photon techniques	Good. Even specific neurons and their components can be targeted

Experimental methods for selective modulation of the activity of particular parts of the brain have been developed for various spatial and temporal scales.

usefulness in specific network types: propagation of spiking activity in feedforward networks is not influenced by the type of neuron model, and feedforward networks with integrate-and-fire neurons [38,39] behave similarly to those with detailed neuron models [40]. Likewise, in a study of the response of different types of neuron models (firing-rate model, conductance-based leaky-integrate-and-fire neurons) to higher-order input activity correlations, the input statistics proved much more important than the neuron type [41].

At the microscopic scale, it is instructive to investigate how a given neuron influences local and downstream network activity, given its other properties (morphology, gene expression, firing pattern, neurotransmitter type) [42]. Recent experiments suggest that neurons indistinguishable by existing classification schemes might form functional subpopulations [43]. Activity-dependent labeling showed that a comparatively small group of neocortical pyramidal cells might provide the major background drive in their respective networks [44]. Similarly, long-range projection targets of cortical output cells are predicted by their intracortical connectivity [45], suggesting that it might be informative to classify them by their afferents and efferents. The density of outgoing projections, together with their overall activity, could be combined into a single variable, quantifying how strongly a cell (cell group) influences the network activity. This, together with the type of neurotransmitter and, to some extent, the synaptic properties, may be used to define the 'embeddedness' of a neuron (Box 1) [42].

Selectively modulating individual neurons and recording the network response allows the 'effective' embeddedness of a neuron *in vivo* to be estimated experimentally. In addition, selectively visualizing the presynaptic sources [46,47] or postsynaptic targets [48] of a neuron may provide useful insights into its 'structural' embeddedness. The structural embeddedness of a neuron in its local microcircuit may also be estimated by juxtacellular recording *in vivo* and labeling afterwards [49]. Next to these indirect methods, new approaches for measuring neuronal embeddedness need to be established. Indeed, optogenetic tools rank among the most promising candidates in this respect.

Controllability as a concept for selective modulation in complex networks

When considering the brain as a dynamical system, firing rates of neurons or average population rates of participating networks are often used as dynamic variables. Traditional approaches such as eigenvalue or Schur decomposition, transforming a network into a new system with only self-interactions or feedforward interactions, respectively, are extremely successful in helping to understand physical systems. However, it has thus far not been possible to map these concepts to experimental studies of neuronal networks [5].

Although it could be held that the ultimate goal of neuroscience is to understand the functional role of each neuronal network and neuron type in the brain, there are fundamental problems with this. We discuss here a new

Box 1. Embeddedness

The concept of embeddedness was initially coined for socio-economic networks to understand the effect of social relations on economic decision-making [64]. Within the context of neuroscience, embeddedness measures the impact of the spiking activity of a neuron on the spiking activity of the surrounding network. It can be experimentally estimated by stimulating a neuron and counting the total number of 'extra spikes' in the network [65] (Figure 4).

When a neuron is activated, it sends spikes to its postsynaptic neurons; some of which will produce spikes and send those to their postsynaptic neurons, and so on. Thus, to estimate the impact of activating a neuron we need to know all possible direct and indirect (involving multiple synapses) paths from the stimulated neuron to other neurons in the network. That is, the sum of the powers of the connectivity matrix ($\Sigma_N A^N$) can be used to estimate the number of extra spikes induced by stimulating a neuron or a group of neurons [66]. When the embeddedness is estimated using the connectivity matrix, we refer to it as 'structural embeddedness'.

Although the connectivity matrix is an important determinant, so far it has not been possible to associate a particular graph property alone with embeddedness because synaptic and cellular properties, ongoing activity, neuromodulators, etc. also shape the impact of a neuron on the dynamics of the network. The experimental estimate of 'extra spikes' implicitly includes all these other factors as well. When the effective connectivity matrix A_{eff} (measured from the neuronal activity) is known, the 'effective embeddedness' could be estimated using the powers of A_{eff}. Finally, graph-theoretical measures such as out-degree, k-shell index, eigenvalue centrality etc. are also related to the embeddedness (Figure 4). For details see [42].

Box 2. Controllability

Consider a network of neurons or networks:

$$\dot{\lambda} = -I\lambda + A\lambda + BU + \xi \quad [I]$$

where λ is the column vector of firing rates of the N neurons (neuronal populations in an NoN), A is the connectivity matrix (size $N \times N$), representing the connectivity among the neurons or the neuronal populations, respectively, I is the identity matrix describing the fact that the activity in a network would change in the absence of inputs from other nodes and external sources, U (size $R \times 1$) represents the R external inputs, B is the connectivity of the neuronal populations with the external inputs (size $N \times R$), and ξ is the noise in each of the neuronal populations.

All activity states of an N-node network can be described in an N-dimensional space (Figure 3). Controllability determines whether a linear system (Equation I) allows a transition from one dynamical state to another in finite time [50]. The $N \times NR$ controllability matrix (C) is defined as:

$$C = [B\, AB\, A^2B \ldots A^{N-1}B] \quad [II]$$

For a fully controllable system, the matrix C should have a full rank, in other words, $Rank(C) = N$. This ensures that all dynamical states of the system are accessible.

Because the exact values of the matrices A and B are not known for most neuronal systems, C cannot be evaluated. However, the notion of controllability can be extended to the concept of structural controllability [51]. To estimate the structural controllability matrix C^S, all non-zero entries in the matrices A and B are replaced with 1 s, reducing the connectivity matrices to binary adjacency matrices [67]. If $Rank(C^S) = N$, the system is 'controllable' [51] and the controllability of a network can be determined, even if the exact values of the connection strengths are not known.

Furthermore, the concepts of 'driver nodes' [52] and 'power dominant set' [53] allow us to calculate the minimal set of nodes that need to be stimulated to control the network dynamics and suffices to visit all activity states of the network.

Driver nodes: the set of the minimum number of nodes to achieve full control over the network dynamics. To identify the driver nodes we need to find the 'maximum matching' in the network, which refers to the maximum set of links that do not share starting or end nodes [52]. A node is matched if a link in the maximum matching set points towards it; otherwise, it is unmatched. The unmatched nodes are the driver nodes that need to be directly stimulated to control the network dynamics. For instance, in Figure 2C the first node P1 is unmatched.

Power dominant set (PDS): when the nodes have self-couplings (Figure 2C, dotted arrows), every node could become a driver node [53]. For such systems the control nodes are the minimum number of nodes from which the rest of the nodes are only one link away [53]. In Figure 2C, the nodes P1, P3, P5 or P1, P2, P4 constitute the PDS.

paradigm based on the concept of controllability (Box 2), which might be a promising candidate to assist in understanding the function and dynamics of neuronal networks. Interestingly, in this paradigm we could potentially avoid the combinatorial explosion. Based on the dynamical description of neuronal network activity (Equation I in Box 2), the function of the brain could be considered as the ability to find desired trajectories of the network activity state (Box 2) because different behavioral states can be mapped onto these activity states. Following this interpretation, a more pertinent question in understanding brain function would be to know how the overall system dynamics could be controlled, rather than to determine how one particular neuron population affects the activity of the others.

However, even in this more 'utilitarian' view, the problem still arises concerning which specific nodes should be stimulated, and how to 'steer' the network to a desired state, even when the connectivity matrix A is known. A trivial, but biologically non-implementable, solution to control the network dynamics is to stimulate all nodes externally. By contrast, a systematic search for the set of nodes to stimulate for controlling the system leads to the afore-described combinatorial explosion.

Engineers have successfully applied the concept of controllability to predict whether a given system can be 'steered' from one dynamical state to another in finite time [50] (Box 2). For a dynamical system, controllability depends on the matrices A and B. Fortunately, in the absence of exact values of A and B, the notion of structural controllability (Box 2) [51] can be effectively used to determine the controllability of a system, and it is possible to identify a minimum set of nodes, the stimulation of which is sufficient to control the full network state [52]. Thus, equipped with the knowledge of A, we can bypass the brute force approach and identify the 'driver nodes' whose stimulation can control the dynamical repertoire of the neuronal network.

The estimation of the driver nodes explicitly assumes that the nodes/neurons do not have self-connections, which is true for most neurons. However, at the system dynamics level the self-connections capture the intrinsic dynamics of the nodes – in other words, node dynamics in the absence of influences from other nodes and neuron refractoriness. Ignoring the intrinsic dynamics implicitly means that the activity state of the node does not change when there are no influences from other nodes (i.e., $\lambda = 0$ or infinite time constant) [53]. Typically, the activity of biological neurons/networks decays to zero in the absence of external inputs. Thus, the absence of self-connections can only be justified in restricted cases, for example, when the dynamics of the constituent networks of a NoN exhibit attractor dynamics and can exhibit asynchronous-irregular self-sustained activity [54]. Furthermore, it has been shown that the 'power dominant set' (PDS) (Box 2) is the set of driver nodes when intrinsic dynamics and self-connections are included in the dynamics [53]. In this context it is important to note that hubs, which are typical choices for external brain stimulation, usually are not part of the driver nodes and PDS [52].

Extending the framework of controllability to neuronal networks, we argue that the concepts of driver nodes [52] and the PDS [53] have emerged as interesting alternatives for choosing the nodes to be stimulated. Once the appropriate set of driver nodes is known, the controllability Gramian [55] can be used to determine the input pattern needed to drive the system to a desired state. Because the number of neurons in a set of driver nodes or PDS is typically substantially less than the number of nodes in the network, the total number of stimulations required will always be significantly less than Bell's number. However, the framework of controllability still requires simultaneous manipulations of multiple nodes. Therefore, we will certainly need technology that is capable of varying the activity of more than few nodes independently. In addition to the notion of controllability, other graph-theoretical measures [56,57] may also be used to make educated choices of the stimulation sites.

Concluding remarks

The considerations outlined above identified two major pitfalls when applying selective activity modulation to understanding brain function: (i) the need for simultaneous control of the activity of multiple neuronal populations leads to a severe combinatorial explosion; (ii) using gene expression, firing patterns, morphological or spatial location criteria alone to define functional groups of neurons is highly ambiguous. Interestingly, in other scientific disciplines it has been realized that the seemingly intuitive approach of selective activity modulation of parts of a complex system to study its function can be misleading. For instance, systems biologists now agree that genes themselves act as a complicated network of dependencies, and that single-gene/single-function mapping is the exception rather than the rule [58]. We argue that the dynamic and nonlinear nature of interactions within large neuronal networks in the mammalian brain makes it unlikely that single-neuron-type/single-function relationships hold for these systems, either. Recent cross-modal studies even suggest that the separation between areas concerning their involvement in different functions may be less clear than previously thought [28–31].

Instead, brain function is more likely to be the result of coordinated activity distributed over multiple brain areas. Assuming that brain function can be understood as an interaction of subnetworks of different types of neurons ignores this insight, as well as the rich repertoire of activity dynamics that can be displayed by neuron subtypes depending on the activity of the embedding network [59]. Finally, the approach of defining cell populations based on currently available markers can only be a starting point and needs to be complemented by more sophisticated criteria. Embeddedness, graph-theoretical measures, and criteria derived from linear and nonlinear control theory [60] may be promising candidates in this respect.

Future directions

We have shown that knowing the network connectivity of neurons and neuronal populations can help in choosing the most appropriate network node(s) for activity modulation to help understand the function and dynamics of networks in the brain. Instead of needing to go painstakingly through all possible combinations, such approaches (e.g., based on controllability) could potentially evade some of the inherent limitations of selective activity modulation approaches. Indeed, recent developments of optogenetic tools provide interesting avenues in this direction. Hence, future research on the development of experimental tools should be driven by the following goals:

(i) To define tools that can be used to control the activity of multiple neuronal populations simultaneously.
(ii) To identify biomarkers for classifying neurons according to their recent spiking activity [43,61].
(iii) To extract the functional and anatomical connectivity of different types of neurons and neuronal populations, and to use these to measure their embeddedness and the controllability of the network [42].
(iv) To design selective activity modulation experiments based on the network-related features extracted with the above-mentioned approaches.

In parallel, it is also important to develop new computational models, where selective activity modulation-based experimental approaches can be tested under fully controlled conditions. As a starting point, we provide an online system (Neural System Prediction and Identification Challenge, nuSPIC) for extracting the function of relatively small networks of spiking neurons designed to perform a specific task. Its web interface provides several tools for performing a variety of experiments, including selective activity modulation. We believe that nuSPIC provides useful models for calibrating the efficacy of experimental approaches and to help interpret their results. Once we have tools to control selectively the activity of multiple neuronal populations simultaneously, and the knowledge of their mutual connectivity, we will be in a position to develop a deeper understanding of their contribution to brain function.

Acknowledgments

We thank Nikos Logothetis, Moshe Abeles, Stefan Rotter, and Yexica Aponte for helpful discussions and comments on the initial version of the manuscript. Partial funding by German Federal Ministry of Education and Research (BMBF) grant 01GQ0420 to BCCN Freiburg, 01GQ0830 to the Bernstein Focus Neurotechnology (BFNT) Freiburg/Tuebingen, and the BrainLinks-BrainTools Cluster of Excellence funded by the German Research Foundation (DFG) grant EXC 1086, is gratefully acknowledged.

References

1 Fellman, S.J. and Van Essen, D.C. (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–46
2 Modha, D.S. and Singh, R. (2010) Network architecture of the long-distance pathways in the macaque brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13485–13490
3 Murayama, M. *et al.* (2009) Dendritic encoding of sensory stimuli controlled by deep cortical interneurons. *Nature* 457, 1137–1141
4 Xu, N.L. *et al.* (2012) Nonlinear dendritic integration of sensory and motor input during an active sensing task. *Nature* 492, 247–251
5 Goldman, M.S. (2009) Memory without feedback in a neural network. *Neuron* 61, 621–634
6 Johannesma, P.I.M. and Aertsen, A. (1987) Conservation and dissipation in neurodynamics. In *Physics of Cognitive Processes* (Caianiello, E.R., ed.), pp. 228–257, World Scientific Publishing
7 Johannesma, P.I.M. *et al.* (1986) From synchrony to harmony: Ideas on the function of neural assemblies and on the interpretation of neural synchrony. In *Brain Theory* (Palm, G. and Aertsen, A., eds), pp. 25–47, Springer
8 Koch, C. (2012) Modular biological complexity. *Science* 337, 531–532
9 Rota, G.-C. (1964) The number of partitions of a set. *Am. Math. Monthly,* 71, 498–504
10 Boucsein, C. *et al.* (2011) Beyond the cortical column: abundance and physiology of horizontal connections imply a strong role for inputs from the surround. *Front. Neurosci.* 5, 32
11 Binzegger, T. *et al.* (2004) A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.* 24, 8441–8453
12 Kremkow, J. *et al.* (2007) Emergence of population synchrony in a layered network of the cat visual cortex. *Neurocomputing* 70, 2069–2073
13 Wagatsuma, N. *et al.* (2011) Layer-dependent attentional processing by top-down signals in a visual cortical microcircuit model. *Front. Comput. Neurosci.* 5, 31
14 Olsen, S.R. *et al.* (2012) Gain control by layer six in cortical circuits of vision. *Nature* 482, 47–52
15 Lee, J.H. *et al.* (2010) Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. *Nature* 465, 788–792
16 Logothetis, N.K. (2010) Bold claims for optogenetics. *Nature* 468, E3–E4
17 Janusonis, S. (2012) Relationships among variables and their equilibrium values: caveats of time-less interpretation. *Biol. Rev. Camb. Philos. Soc.* 87, 275–289

18 Mehring, C. *et al.* (2003) Inference of hand movements from local field potentials in monkey motor cortex. *Nat. Neurosci.* 6, 1253–1254
19 Nunez, P. (2006) *Electric Fields of the Brain: the Neurophysics of EEG*. Oxford University Press
20 Anastassiou, C.A. *et al.* (2011) Ephaptic coupling of cortical neurons. *Nat. Neurosci.* 14, 217–223
21 Fröhlich, F. and Mccormick, D.A. (2010) Endogenous electric fields may guide neocortical network activity. *Neuron* 67, 129–143
22 Luria, A.R. (1973) *The Working Brain: An Introduction to Neuropsychology*. Allen Lane
23 Fritsch, G. and Hitzig, E. (1870) Uber die elektrische Erregbarkeit des Grosshirns. *Arch. Anat. Physiol.* 37, 300–332
24 Logothetis, N.K. *et al.* (2010) The effects of electrical microstimulation on cortical signal propagation. *Nat. Neurosci.* 13, 1283–1291
25 Boyden, E.S. *et al.* (2005) Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* 8, 1263–1268
26 Miesenböck, G. and Kevrekidis, I.G. (2005) Optical imaging and control of genetically designated neurons in functioning circuits. *Annu. Rev. Neurosci.* 28, 533–563
27 Cisek, P. and Kalaska, J.F. (2010) Neural mechanisms for Interacting with a world full of action choices. *Annu. Rev. Neurosci.* 33, 269–298
28 Bizley, J.K. *et al.* (2007) Physiological and anatomical evidence for multisensory interactions in auditory cortex. *Cereb. Cortex* 17, 2172–2189
29 Ghazanfar, A.A. *et al.* (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J. Neurosci.* 25, 5004–5012
30 Iurilli, G. *et al.* (2012) Sound-driven synaptic inhibition in primary visual cortex. *Neuron* 73, 814–828
31 Kayser, C. *et al.* (2008) Visual modulation of neurons in auditory cortex. *Cereb. Cortex* 18, 1560–1574
32 Horton, J.C. and Adams, D.L. (2005) The cortical column: a structure without a function. *Philos. Trans. R. Soc. Lond. Ser. B: Biol. Sci.* 360, 837–862
33 Kremkow, J. *et al.* (2010) Functional consequences of correlated excitatory and inhibitory conductances in cortical networks. *J. Comput. Neurosci.* 28, 579–594
34 Freund, T.F. and Buzsáki, G. (1996) Interneurons of the hippocampus. *Hippocampus* 6, 347–470
35 Luo, L. *et al.* (2008) Genetic dissection of neural circuits. *Neuron* 57, 634–660
36 Markram, H. *et al.* (2004) Interneurons of the neocortical inhibitory system. *Nat. Rev. Neurosci.* 5, 793–807
37 Migliore, M. and Shepherd, G.M. (2005) An integrated approach to classifying neuronal phenotypes. *Nat. Rev. Neurosci.* 6, 810–818
38 Diesmann, M. *et al.* (1999) Stable propagation of synchronous spiking in cortical neural networks. *Nature* 402, 529–533
39 Kumar, A. *et al.* (2010) Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nat. Rev. Neurosci.* 11, 615–627
40 Shinozaki, T. *et al.* (2007) Controlling synfire chain by inhibitory synaptic input. *J. Phys. Soc. Jpn.* 76, 044806
41 Kuhn, A. *et al.* (2003) Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Comput.* 15, 67–101
42 Vlachos, I. *et al.* (2012) Beyond statistical significance: implications of network structure on neuronal activity. *PLoS Comput. Biol.* 8, e1002311
43 Liu, X. *et al.* (2012) Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* 484, 381–385
44 Yassin, L. *et al.* (2010) An embedded subnetwork of highly active neurons in the neocortex. *Neuron* 68, 1043–1050
45 Brown, S.P. and Hestrin, S. (2009) Intracortical circuits of pyramidal neurons reflect their long-range axonal targets. *Nature* 457, 1133–1136
46 Bock, D.D. *et al.* (2011) Network anatomy and in vivo physiology of visual cortical neurons. *Nature* 471, 177–182
47 Rancz, E.A. *et al.* (2011) Transfection via whole-cell recording in vivo: bridging single-cell physiology, genetics and connectomics. *Nat. Neurosci.* 14, 527–532
48 Komiyama, T. *et al.* (2010) Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. *Nature* 464, 1182–1186
49 Burgalossi, A. *et al.* (2011) Microcircuits of functionally identified neurons in the rat medial entorhinal cortex. *Neuron* 70, 773–786
50 Kalman, R.E. (1963) Mathematical description of linear dynamical systems. *J. Soc. Ind. Appl. Math. Ser. A* 1, 152–192
51 Lin, C-T. (1974) Structural controllability. *IEEE Trans. Automatic Control* 19, 201–208
52 Liu, Y-Y. *et al.* (2011) Controllability of complex networks. *Nature* 473, 167–173
53 Cowan, N.J. *et al.* (2012) Nodal dynamics, not degree distributions, determine the structural controllability of complex networks. *PLoS ONE* 7, e38398
54 Kumar, A. *et al.* (2008) The high-conductance state of cortical networks. *Neural Computation* 20, 1–43
55 Rugh, W.J. (1996) *Linear System Theory* (2nd edn), Prentice-Hall
56 Dorogovtsev, S.N. and Mendes, J.F.F. (2003) *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press
57 Newman, M.E.J. (2003) The structure and function of complex networks. *SIAM Rev.* 45, 167–256
58 Chouard, T. (2008) Beneath the surface. *Nature* 456, 300–303
59 Wang, X-J. (2010) Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol. Rev.* 90, 1195–1268
60 Slotine, J-J. and Li, W. (1991) *Applied Nonlinear Control*. Prentice Hall
61 Garner, A.R. *et al.* (2012) Generation of a synthetic memory trace. *Science* 335, 1513–1516
62 Ranck, J.B. (1975) Which elements are excited in electrical stimulation of mammalian central nervous system: a review. *Brain Res.* 98, 417–440
63 Yizhar, O. *et al.* (2011) Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* 477, 171–178
64 Granovetter, M. (1985) Economic action and social structure: the problem of embeddedness. *Am. J. Sociol.* 91, 481–510
65 London, M. *et al.* (2010) Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature* 466, 123–127
66 Pernice, V. *et al.* (2011) How structure determines correlations in neuronal networks. *PLoS Comp. Biol.* 7, e1002059
67 Newman, M.E.J. (2010) *Graph Theory and Complex Networks: An Introduction*. Maarten van Steen