

Modeling Coordination in the Neocortex at the Microcircuit and Global Network Level

Anders Lansner and Mikael Lundqvist

Abstract

A key role for computational modeling in neuroscience is to connect cortical microscopic processes at the cellular and synaptic level with large-scale cortical dynamics and coordination underlying perceptual and cognitive functions. Data-driven and hypothesis-driven approaches have complementary roles in such modeling. The Hebbian cell assembly and attractor network paradigm has a potential to explain the holistic processing and global coordination characteristic of cortical information processing and dynamics. The pros and cons of such a view are described. A large-scale model of cortical layers 2/3 formulated along these lines exhibit the fundamental holistic perceptual and associative memory functions performed by the cortex. Such a model can provide important insights into the possible roles of oscillations and synchrony for processing and dynamic coordination and it highlights important issues related to cortical connectivity, modularization and layered structure.

Introduction

Today, massive amounts of data are available about the brain from many different sources as well as from molecular, subcellular, neuronal, and network levels, and more keeps accumulating due to increasingly advanced measurement techniques. Despite this, our mechanistic understanding of the normal and dysfunctional brain, in terms of how processes at these different levels interact dynamically to produce cognitive phenomena and overt behavior, is greatly lacking. Such an understanding would clearly open up new avenues in the search for more effective drugs and therapies for severe diseases and disorders affecting the brain. Perhaps the only tools that offer some hope of eventually reaching this understanding are mathematical modeling and computer simulation. A computational model can organize efficiently new data in a

coherent fashion so as to generate new experimentally testable questions. This may eventually enable us to formulate quantitative theoretical models and thus achieve a more general understanding of the phenomena under study.

By necessity, the acquisition of data about the brain is largely determined by what is possible to measure, not by what we need to know to build better computational models. This tends to cause some detachment between experimentalists and modelers in the sense that the latter experience a patchiness of data where critical pieces are missing, whereas the former feel they produce data that is ignored by modelers. In the early stages of formalizing a scientific field, the approach has typically been one of starting from the fundamentals and gradually adding more detail and complexity. In some instances of less complex and more accessible neural systems, we have already seen a productive interaction between experimental and computational neuroscience (Kozlov et al. 2009), and we expect this to develop further in the near future.

As in other fields, multiscale modeling will be important in brain science, as it will allow us to relate detailed dynamical processes at the cellular, subcellular, and microcircuit level with cognitive phenomena at the level of brain-scale neuronal networks. Software tools, which allow large-scale network models comprising a mix of biophysically and biochemically detailed spiking neuron models, simplified integrate-and-fire neurons, and mesoscopic nonspiking population units to be defined and simulated, have been developed and are now available in the field of neuroinformatics. The most simplified mesoscopic models may represent some cortical regions with only one or two units and thus connect to, for example, dynamic causal modeling techniques used for analysis of brain imaging data.

The neocortex, which is the largest part of the human brain and the site of higher cognitive functions, has been a favorite subject for brain modelers. Computational models with some level of biological detail have been proposed and investigated for more than half a century. As more data has been acquired and as computers have become more powerful, models have increased in size and sophistication. Today, supercomputers are used to model large-scale neuronal networks with a high degree of detail in their component neurons and synapses. If current trends in computing continue, we should be able to simulate in real time detailed models of the entire human brain in about fifteen to twenty years.

Since there has been and still is quite some uncertainty with respect to the relevant experimental data, every neuronal network model to date is explorative and hypothetical. Nevertheless, several of even the early cortex models have been able to display interesting features of, for example, associative memory, dynamic activity, and coordination, replicating some of the key characteristics of the system modeled. Although abstract and without much biological detail, these models nevertheless provided a framework for studying important phenomena beyond what could be grasped intuitively and conceptually. Only by continuing in this same spirit and by bringing models closer to

data can we maintain some hope of reaching a true mechanistic understanding of this very complex system.

In this chapter, we focus on attempts in computational neuroscience to connect cortical microscopic processes, at the cellular and synaptic level, with large-scale cortical dynamics and coordination more directly related to perceptual and cognitive functions. We discuss to what extent models might be data driven and the role of hypotheses in this research, as well as possible interpretations and functional implications of the dynamics observed in large-scale cortex simulations. We consider the role of oscillations and synchrony for processing and dynamic coordination. Finally, we argue for complementing the now dominating reductionist approach with more of a synthesis of components to achieve a coherent global picture. This will allow us to discuss the new directions in brain modeling, with more of model-driven experiments.

Data- and Hypotheses-driven Approaches to Modeling

Despite the enormous efforts in experimental brain science, one of the primary challenges of brain modeling is the relative patchiness of data. There is indeed a lot of data available, but this abundance has been driven largely by the capabilities of our recording equipment, rather than by what is the most important for building quantitative models. Heroic attempts to build “bottom-up” data-driven models of a part of the neocortical network without additional assumptions or hypotheses about function and mechanisms are therefore very high risk undertakings. It may be possible to gather high resolution data about single cells and synapses, as well as population data about types of cells and synapses and the distribution and intricate function of ion channels and neuro-modulators acting on them. However, modeling work shows that, for instance, the dynamic function of and information processing performed by a cortical neuronal network is likely to be critically dependent on its synaptic connectivity at a local and global scale (Figure 6.1). The cortical networks are to a significant extent formed by activity-dependent processes individual for each brain. Therefore, with current technology it will be hard, if not impossible, to get the wiring right. Statistical approaches and pooling of data from different animals and regions is most likely to miss their target. Likewise, detailed reconstructions of local pieces of cortex are still very limited in their spatial extent. Nevertheless, such projects will provide valuable information to constrain and refine current and future quantitative and computational models.

An alternative to modeling is a hypothesis-driven or “top-down” approach. Starting from some conceptual theory of the fundamental functions of cortex, one builds a model gradually, by entering the basic elements first and adding additional relevant details one by one. This strategy has historically proven quite successful in many scientific and engineering fields. A useful theory should take into account data from many different sources and levels of description:

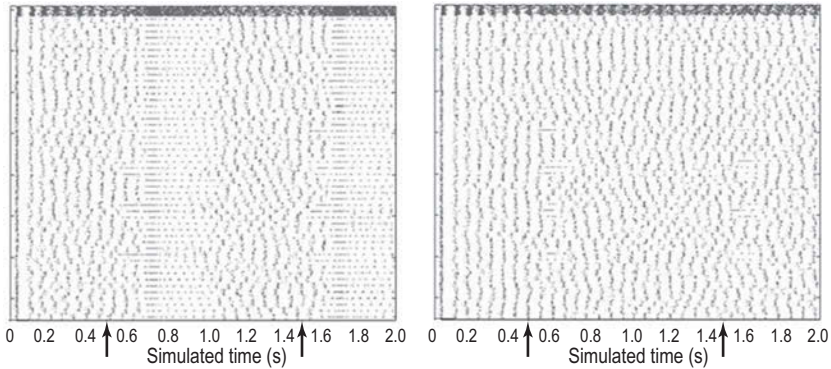


Figure 6.1 Dynamics of network with structured vs. permuted connection matrix. The figure shows the spike raster from runs of two networks. Patterned stimuli are given at the times indicated by the arrows. They differed only in that the left one had memory patterns stored in attractors; the right one had the same connections as the left, except that the pre- and postsynaptic cells were randomly permuted. Thus, both network connectivities obeyed the experimentally measured pair-wise connection statistics between different neuron types. As can be seen, this difference results in dramatically different responses to a patterned input.

from anatomical and physiological data on the neuronal and synaptic levels to experimental observations at the psychological and cognitive levels. Synaptic connectivity may be formed by using a Hebbian learning rule to store a set of memory patterns. Current hypothesis-driven models of cortex take only a fraction of the data available into account and thus are always at risk of leaving out relevant information. Typically, modelers need to fill in some critical data that is missing or conflicting. These parts of the model should be regarded as predictions. If later falsified by new information, models must be reexamined and, if no work-around is found, they should be revised or entirely discarded.

Hebbian Cell Assemblies and the Attractor Network Paradigm

Though the diversity of hypotheses about basic principles of operation of cortex has been pruned over the years, through exposure to increasing amounts of experimental data, a number of different ones still remain. Among some of the more vital and general theories of cortical function today, we find synfire chains (Abeles 1991), liquid state machines (Maass et al. 2002), and localist models of cognition (Bowers 2009). However, the oldest and currently most well-developed theory dates back to Hebb's cell assembly theory (1949). It has more recently been formalized in terms of attractor networks (Hopfield 1982; Amit 1989) and has been extensively studied, both mathematically and by means of computer simulations.

The most well-known component of Hebb's theory is that of coactivation-triggered (Hebbian) synaptic plasticity, and this part has been largely verified by later experiments, which have also extended our knowledge about synaptic plasticity far beyond the original proposal. The more controversial and less acknowledged aspects of Hebb's theory concern the network-level consequences of such processes. Hebb proposed that the wiring together of groups of neurons activated by the same stimulus would form mental objects in the cortex, which he called cell assemblies. They constitute the internal representations of the corresponding objects in the external world, and their existence changes the network dynamics, resulting in phenomena such as after-activity, figure-ground segmentation, and perceptual completion and rivalry effects. This view allows fundamental cognitive functions (e.g., content-addressable associative memory and association chains) to be conceptually understood in terms of processes at the neuronal and synaptic level. Several of these properties are prominent in the corresponding abstract computational models (e.g., a Hopfield network model of associative memory).

The term "attractor" stems from the analysis of the dynamical properties of such abstract recurrent artificial neural networks. Due to the symmetric connection matrix, their state always evolves from an initial state to the closest fix-point attractor and remains there. If a slow dynamics, corresponding to neuronal adaptation or synaptic depression, is introduced in such a model, the attractor state is destabilized and may shift its position in state space or even become transient; that is, it is not formally a fix-point attractor any longer. Such a network typically displays a more complex limit cycle dynamics, which can be described as the state jumping between the memorized states, visiting them for some time determined by the time constants of adaptation and depression, typically in the order of some hundreds of milliseconds. We refer to this here as the attractor states corresponding to memorized patterns becoming quasi-stable.

Adaptation and synaptic depression are common features of, for example, pyramidal cells and synapses between these in layers 2/3. Therefore, a similar dynamics as described above, but even more complex due to the spiking process itself, can be seen in networks of biophysically detailed model neurons and synapses. Here the state of a unit in the abstract network corresponds to, for example, the average instantaneous spiking frequency of the neurons in a local population of some tens or hundreds of neurons.

It needs to be emphasized that the basic attractor network paradigm primarily takes into account the recurrent cortical connectivity most prominent in layers 2/3 and 5. The holistic processing is assumed to be supported by this connectivity and the more refined, specific, and invariant neuronal response properties and representations likely found in higher-order sensory and association cortex. This view is thus incomplete as it disregards the important aspects of how such response properties and representations are formed by self-organization and learning. Such processes can be attributed to the feedforward

processing stream of cortex with layer 4 as a key player in transforming the input from different sensor arrays to a sparse, decorrelated, and distributed code suitable for further attractor network processing. In the following, we will assume that the internal representations have already been transformed in this manner.

The cortical network, with its feedforward, lateral, and feedback projections, forms a brain-scale recurrent network structure which may support global attractor dynamics and top-down influences on earlier sensory areas. Attractors extending over large parts of cortex would exert a powerful dynamic coordination (e.g., in the form of multimodal integration). A current target of large-scale cortex simulations is to find out if the known strength and distribution of cortical long-range connectivity is sufficient to sustain such wide area coordinated and quasi-stable attractor states.

Criticisms of the Attractor Network Hypothesis

Though attractive as a working hypothesis for the primary holistic “Gestalt” processing functions of the cortex, this paradigm has been criticized on several points. First, it is obvious that the abstract models have very simplistic units and connectivity, and that they violate Dale’s law of separate excitatory and inhibitory units. Further, their dense activity and full connectivity is quite different from the sparse activity and highly diluted connectivity typical for real cortex. In addition, the connection matrix needs to be symmetric to guarantee a well-behaved dynamics, and this seems highly unlikely to hold in reality. A spiking attractor network was also expected to have problems generating the low rate irregular discharge patterns of neurons *in vivo*, and a network with neurons firing at low rate was suspected to converge too slowly. Here we will discuss these criticisms and show how many of them have now been shown to be invalid. We will illustrate how attractor memory models have been brought closer to biological reality in terms of their constituent components and network structure as well as the kind of attractor dynamics and oscillatory activity they display.

There were further concerns about the storage capacity of the original Hopfield network being unreasonably low. However, later theoretical analysis and simulations have shown that an attractor network is actually an efficient associative memory in an information theoretic sense, and that its storage capacity scales linearly with the number of synapses (Amit et al. 1987). With realistically low activity levels, the number of distinct attractors possible to store is much larger than the number of units in the network.

Let us now turn to the biological plausibility of this paradigm and describe a plausible mapping of the attractor network paradigm to a simulation model of cortical layers 2/3. We have used this model to interpret experimental data

on perceptual processing and associative memory functions as well as cortical oscillatory phenomena and synchrony at different frequencies and timescales.

Cortical Local Subnetworks and Functional Microcircuits

When building a cortex model, one must confront, early on, the issue of possible repetitive elements other than neurons. It has repeatedly been suggested and discussed that the cortex actually comprises a mosaic of modules, such as functional columns or other types of subnetworks (Mountcastle 1978; Favorov and Kelly 1994; Rockland and Ichinohe 2004; Yoshimura et al. 2005). Such local networks would be more densely connected (like 10–25%) within themselves than to the outside and would likely be selectively targeted by incoming fibers from thalamus. They may be spatially segregated in, for instance, an anatomical minicolumn or may be anatomically diffuse (i.e., intermingled with other similar modules). Possibly in smaller brains (e.g., in the mouse), single or only a few neurons would be in each module; however, in larger cortices (e.g., in primates) each module may need to be more connected than a single pyramidal cell could reasonably sustain. This would make it necessary for several pyramidal cells to cooperate, such that the modules would be larger and perhaps even more anatomically distinct (DeFelipe 2006). Such a minicolumnar structure might also generate a patchy long-range connectivity, as is seen in cortex (Fitzpatrick 1996).

Other data suggests that these local subnetworks are organized in bundles to form larger modules (i.e., macrocolumns, hypercolumns, or barrels; Hubel and Wiesel 1977) and there are abstract models that incorporate such a modular structure (Kanter 1988; Sandberg et al. 2003). A hypercolumn may be assumed to represent, in a discretely coded fashion, some attribute of the external world. For instance, in the primary visual cortex the orientation or direction of an edge stimulus at a certain position on the retina is assumed to be represented by elevated activity in a corresponding orientation column in primary visual cortex. This also leads to sparse activity in the network, on the order of about 1–5%. Such an activity level is in accordance with overall activity densities of about 1% and an average spiking frequency of 0.16 Hz, estimated from metabolic constraints (Lennie 2003).

To investigate the extent to which the attractor network paradigm is compatible with this data, we have designed and simulated a biophysically detailed large-scale model of cortical layers 2/3 (Lundqvist et al. 2006). In this top-down model, the microcircuitry implements two types of modular structures, corresponding to minicolumns and hypercolumns. Each minicolumn is a network unit and comprises thirty pyramidal cells and a couple of dendritic targeting and vertically projecting interneurons, for example, double bouquet (RSNP) cells that inhibit cells in the minicolumn to which they belong. Their role is described in the next section. Each hypercolumn is a bundle of up to a hundred minicolumns together with about an equal number of basket cells.

The latter are driven by nearby pyramidal cells and their axonal arborizations extend over the entire hypercolumn, providing negative feedback and turning the hypercolumn into a kind of winner-take-all module.

Global Recurrent Connectivity Supporting Holistic Processing

Once this model of the local modular structure of the cortex is in place, it is quite obvious to see how the global, long-range connectivity could implement the memory-encoding connection matrix of the attractor network. The largest-scale model that we have simulated comprised several thousand of such hypercolumn modules, where memory patterns were stored by connecting sets of coactive minicolumns, one per hypercolumn. Hebbian plasticity rules typically give a connection matrix with a positive and a negative component. Our network represented the positive elements of the connection matrix by direct intracortical or corticocortical pyramidal-pyramidal synapses, whereas the negative elements of the connection matrix were disinaptic via double bouquet cells in the target minicolumn. This suggests a specific role and synaptic input for this type of cells, which is compatible with known data but where the details are not yet confirmed. Given some spatial modularization of the target area, in terms of functional minicolumns, this will give rise to the patchy long-range connectivity experimentally observed in several cortical areas.

In our model, the long-range connectivity is very sparse at the cell-to-cell level. Typically, in our largest network, out of the 900 possible long-range excitatory synapses from a source to a target minicolumn, only five are instantiated randomly. This gives a connection matrix which is sparse and nonsymmetric at the microscopic level, but symmetric at the macroscopic level of minicolumn units. As mentioned above, in theory, attractor networks need to have a symmetric connection matrix to have a guaranteed convergence to an attractor. In practice, we find that the network state does converge despite the sparse connectivity and this microscopic deviation from symmetry.

In the largest-scale model, comprising 22 million model neurons and 11 billion synapses, the connection densities and PSP amplitudes have been set to reflect experimental data from, for example, Thompson et al. (2002) and Binzegger et al. (2007). The average number of synapses onto single cells of different kinds are then held fixed as the network is scaled (Djurfeldt et al. 2008), whereby a single cell receives approximately the same synaptic input current regardless of network size. The network is driven by input from a few layer 4 pyramidal cells that are activated directly via noise injection and connects with a probability of 50% to the above pyramidal cells in the same minicolumn.

The sparse and low rate activity in the layer 2/3 network has some important implications for the balance between thalamic input, local connectivity, and long-range contextual influence. Though the latter connections are more numerous, only a few percent will be active due to the sparse activity in the

presynaptic population. Despite their relatively low numbers, thalamic input influences the network activity significantly in the model and can by itself switch the activity from one active state to another.

Attractor Memory Dynamics and Holistic Perceptual Processing

What can we say about the dynamics of memory retrieval in a biologically detailed cortical network model, such as the one described above? The most important questions concern holistic processing in terms of attractor convergence, after-activity, associative memory, and perceptual completion and rivalry. The two latter are clear examples of how long-range “contextual” and coordinating interactions can modify and even override (rivalry) local processing. This aspect has been examined in numerous simulations; all of these operations were found to be prominent, occurring on the same timescale as observed psychophysically, in less than a hundred milliseconds (Lansner and Fransén 1992; Fransén and Lansner 1998; Sommer and Wennekers 2001). This is true even if a fragmented and noisy pattern is given as input or if two patterns are stimulated simultaneously, as in a pattern rivalry situation. The capacity to store and reliably retrieve patterns parallels that of the isomorphic abstract associative memory network.

Due to adaptation and synaptic depression in the network, after-activity will typically be restricted to a few hundred milliseconds, following which the activity terminates. If the input is still on, the second-best matching pattern may activate, and there may be a slow oscillating activity with the two memory states alternately in the foreground. This is reminiscent of the alternating perception of ambiguous stimuli as in the Necker Cube illusion. In other situations (e.g., hippocampal place fields), attractors may be overlapping and temporally chained, thus forming more of a line attractor structure (Lisman and Buzsáki 2008).

Notably, the network dynamics is highly sensitive to the higher-order connectivity structure. The statistics provided by pair-wise recordings or from morphological reconstruction is not enough to determine uniquely the network structure or its dynamics. For instance, the prominent attractor dynamics seen in our model, which results from a long-range connectivity set up according to a Hebbian learning rule to store a number of distributed patterns, is not seen in the same network where the long-range connectivity is permuted randomly (Figure 6.1). In both cases, pair-wise connectivity statistics is constrained in the same way according to experimental data; however, in one, an attractor state is entered whereas in the other only a weak disturbance of the ground state results.

With sufficient unspecific background activation, the state of our simulated network will under certain conditions spontaneously jump between attractors in a complex, presumably chaotic fashion and at about theta frequency (~5 Hz). This may relate to the dynamics of different kinds of ongoing activity

recorded in cortex *in vitro* and in slices *in vivo* (Lehmann et al. 1998; Grinvald et al. 2003; Ikegaya et al. 2004). During such activity, the internal state of a cell may show plateaus and hyperpolarized periods. If the network is stimulated during ongoing spontaneous activity, it typically switches quickly to the stimulated pattern. This switch is, however, somewhat less likely to occur during the early part of a newly triggered attractor state, which has been suggested to provide a mechanistic explanation for the cognitive bottleneck phenomenon of attentional blink (Lundqvist et al. 2006).

The duration and intensity of an attractor state is quite sensitive to changes in model parameters relating to Ca–KCa channels and dynamics (Lundqvist et al. 2006). Several of the monoamine neuromodulators (e.g., serotonin and acetylcholine) act on these processes. This suggests a possible connection between these parameters of the model and neuromodulation of ongoing and stimulus triggered cortical dynamics.

Coordination in the form of burst synchronization at about theta frequency occurs over long distances (with conduction delays of tens of milliseconds) in the model, and it could be assumed to extend over large parts of the cortex. Such coordination would be important, for example, for cross-modal integration and top-down expectancy phenomena and could be observed in local field potential (LFP) and EEG recordings (Engel et al. 2001). Top-down excitation via back-projections from attractors formed in higher cortical areas, as well as facilitatory influences from subcortical structures involved in motivational and drive regulation, may guide attention (Ardid et al. 2007). At the global cortical level, a single coherent activity mode seems to dominate the entire network most of the time due to lateral inhibition. Such competition may ultimately be a mechanism responsible for the apparent unity of our consciousness and the serial nature of thought processes at the macroscopic level, in contrast to the massively parallel processing at the microscopic level.

Microcircuits and Fast Oscillations

Network dynamics, with its quasi-stable attractor states, has the potential to generate a slow rhythmic activity in the theta frequency range. This can be seen even in the mesoscopic models with nonspiking population units. While implementing this kind of network with spiking units, faster rhythmic dynamics emerges in the higher frequency ranges of alpha, beta, and gamma oscillations. The latter has gained special interest due to its correlation with attention, perceptual processing, and consciousness (Engel et al. 2001). Alpha and beta oscillations characteristic of resting conditions are replaced by faster oscillations during active processing. The general trend in experimental data seems to be that faster oscillations are more localized to layers or columns and more short-lived (Sirota et al. 2008). This suggests that the origin of fast (e.g.,

gamma) oscillations may be the microcircuit whereas slower oscillations arise from interactions over longer ranges.

Oscillations in the gamma frequency range emerge readily in simulated microcircuits, as a result of strong feedback inhibition, and can even do so in a network of recurrently connected inhibitory neurons when subject to some background excitation. Notably, gap junctions between inhibitory interneurons are not necessary to generate prominent fast oscillations. The main factors that decide the frequency in excitatory-inhibitory (E-I) networks is the type of connectivity, the relation between excitation and inhibition, and the synaptic time constants (Brunel and Wang 2003).

In our own large-scale simulations (Djurfeldt et al. 2008), global alpha and beta frequency oscillations during inactivity were replaced by gamma frequencies and a more focal pattern of activation (Figure 6.2). Under certain conditions with low background activation, the network model displayed a nonspecific, noncoding ground state in which beta frequency oscillations were prominent (Figure 6.1), and the pyramidal cells fired at around 1 Hz. Stimulus-triggered activation of a coding state increased significantly the power in the gamma band (Figure 6.3).

Gamma oscillations emerged due to the interplay between pyramidal cells of an active minicolumn and feedback inhibition from nearby basket cells. This regime also allowed for an approximate balance between excitatory and

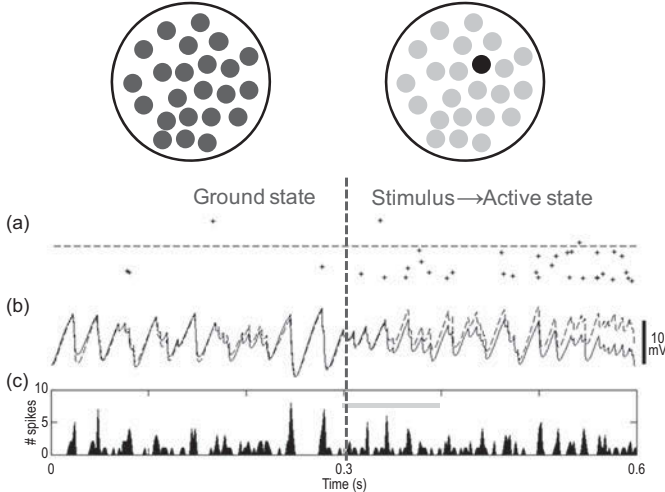


Figure 6.2 Oscillatory activity in ground and active states. Distribution of spiking within a hypercolumn during ground state (left) and active state (right). Average spiking frequency is coded by gray scale for each minicolumn: dark gray < 1 Hz; light gray ~ 2 Hz; black > 5 Hz. (a) Spike raster for the background and foreground minicolumns separated by a dashed line. (b) Average V_m for a foreground (dashed) and background (solid) minicolumn during ground state (0–0.3 s) and active state (0.3–0.6 s). (c) Time histogram of $n:o$ spikes in the entire hypercolumn.

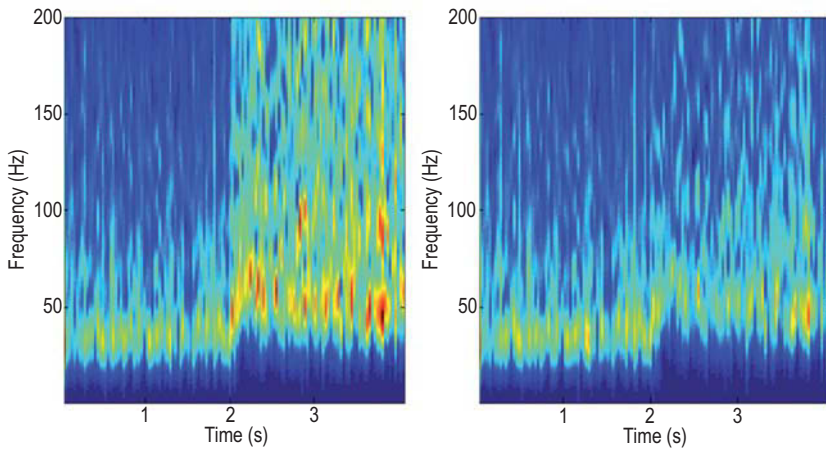


Figure 6.3 Synthetic LFP spectrogram. The network started out in the ground state and entered an active state after two seconds due to stimulation. The signal was produced from 30 local pyramidal cells entering foreground (a) and background (b). The average signal from five runs is plotted.

inhibitory currents in both ground and active states, and as a result cells fired irregularly as seen *in vivo*. In contrast to gamma oscillations, those in the alpha and beta frequency range were generated by activity and feedback inhibition of a much more diffuse and low-rate nature. In the ground state of the network, single cell-firing rates were very low, and the increase in oscillatory frequency thus signaled an increase in selectivity during attractor states.

Furthermore, the increase of cortical gamma oscillations can be entrained by the slower theta rhythm in hippocampus (Sirota et al. 2008). An interpretation of this from the attractor paradigm standpoint would be that (a) theta rhythm results from wandering between task relevant attractors and (b) gamma oscillations can be generated in each attractor state. Thus, according to our simulations, oscillations in several different frequency bands emerge from the underlying attractor dynamics during spontaneous as well as stimulus-driven activity. The slower oscillations seem to reflect the very important mechanism of attractor destabilization which paces spontaneous network activity and the “flow of thought,” but what about the functional role of gamma oscillations?

Does Synchrony Hold Information or Have Computational Advantages?

Due to the local feedback inhibition via basket cells, adjacent pyramidal cells tend to be synchronized in simulated E-I networks. Most network models simulate only this local part of a neural network so that the same inhibitory network of basket cells controls the activity in the entire network. However, if

larger networks that extend beyond the lateral extension of basket cells were investigated, they would involve several hypercolumns. If each of these code for a distinct feature, with minicolumns discretely coding for specific values, would the phase of the gamma oscillations signal which feature values belong together in a group, binding the different parts of an object together globally? If so, synchronization over long distance might help, for instance, to discriminate figure–background, to bind several segments of an object as a single perceived entity, or to form the perception of a novel object by temporary synchronization of previously learned features. Experimental observations support the idea that synchronization may be involved in such binding; for example, if attention is shifted toward a stimulus that is processed by the recorded cells, there is an increase in local coherence in gamma oscillations (Fries, Reynolds et al. 2001). True phase coding should, however, imply synchronization on the millisecond level spanning over several hypercolumns.

One model that displayed feature binding through phase locking was that of Ritz et al. (1994). They showed that a semi-realistic network model could segment a complex scene into previously learned objects by such a mechanism. The representations of different objects oscillated out of phase relative to each other so that temporal synchronization on a fine timescale signaled how features (pixels in this case) were grouped into objects. When the network oscillated at gamma frequencies, up to four objects could be separated. Each population, however, was coupled to distinct interneurons; thus inhibition did not induce competition. Furthermore, the network was not supporting persistent activity and was strictly input driven. Such network architecture is hard to reconcile with what is currently known about cortical connectivity.

Simulations highlight two problems that must be solved for phase coding to work in a realistic cortical network model. First, how can several local populations of pyramidal cells be active simultaneously within about 25 ms if they share the same local inhibitory network? In oscillatory networks pyramidal cell firing is a self-terminating process since it activates strong feedback inhibition. Therefore each oscillatory cycle will set up a racing condition or winner-take-all process, where the phase-leading population will terminate the activity of the other populations (Fries et al. 2007). This is presumably an important mechanism behind the rivalry observed in perceptual processing. According to our experience, it is very hard, if not impossible, to tune a network model with recurrent inhibition so that more than one population fires in each gamma cycle. To resolve this issue it is important to investigate if nearby cortical pyramidal cells receive feedback inhibition from distinct inhibitory networks, or if selective inhibition, via dendritic targeting inhibitory interneurons, is predominantly of a feedforward kind, activated from distant pyramidal cells as in our network model described above.

Second, we need to understand how populations of pyramidal cells that do not share the same local inhibition (i.e., belong to distinct functional hypercolumns) are able to fire phase-locked with each other. Since inhibitory

interneurons are key players in generating locally synchronous oscillatory activity, how is the pace and phase kept over distances longer than those directly accessible to them? One solution might be long-range excitation; however, in our own model, this induces only a weak synchrony between minicolumns in different hypercolumns (Lundqvist et al. 2006). Recent modeling has shown further that synchrony between populations can only be achieved with strong inter-regional coupling (Honey et al. 2007) and that increase in synchrony cannot be separated from an increase in firing rates. This suggests that phase locking does not occur in the absence of rate modulation and that oscillations and synchronization are emergent features of spiking attractor dynamics in networks controlled by strong feedback inhibition.

Thus, it appears to us unlikely that several gamma oscillation modes could robustly exist independently in the same cortical location, or that assemblies are bound together by phase locking over longer distances. Still, oscillations in the gamma range are obviously prominent features of active cortical processing, and they readily emerge in network models. They might still convey computational advantages since the saliency of synchronized signals is heavily increased, allowing for self-sustained activity with very low (metabolically favorable) firing rates. Furthermore, transitions between coding states are very fast in the oscillatory regime (Fries et al. 2007), since there are only small differences in net excitation between foreground and background states, and relatively few thalamic inputs can strongly influence the network activity as described above.

Plasticity in the Microcircuit and Global Network

Activity-dependent plasticity has an obvious role in the attractor network paradigm; namely, in the formation of global cortical connectivity that establishes the attractor structure of the network. A standard form of Hebbian learning seems to be sufficient to support the holistic processing of such networks. However, several interesting issues open up when viewed from a broader perspective.

For example, an STDP-type learning rule seems well suited for a gamma oscillating network and might even impose a temporal structure reminiscent of synfire chains to the attractor dynamics, where subgroups of the assembly are sequentially activated. True sequence learning, as already suggested by Hebb, could readily be achieved in an attractor network provided that the pre-post timing window is somewhat widened and possibly made asymmetric. This would suggest that the synapses involved in global cortical connectivity could display a spectrum of temporal selectivity for LTP induction.

In addition, we note that our depiction is to a significant degree incompatible with attractor models of working memory, based on persistent activity (Camperi and Wang 1998). These models rely on attractor states for actual

storage of memory items over time. Since an attractor network model supports only reluctantly the simultaneous activation of more than one state, multiple-item working memory would pose a problem for a realistic model. Alternatively, we suggest that working memory is based instead on fast-inducing, volatile Hebbian synaptic plasticity (Sandberg et al. 2003), with possible spontaneous cyclic activation of a limited number of attractor states representing items in working memories. Such a mechanism would unify long- and short-term forms of memory, and data on fast synaptic plasticity suggest that this is a realistic possibility (Hoffman et al. 2002).

Furthermore, the type of learning rules expected to operate should be quite different, depending on what type of processing and cortical layer we consider. From the perspective of the attractor network paradigm, the feedforward processing stream via layer 4 is likely to be involved in transforming internal representations to make them suitable for processing in higher-order attractor networks. Decorrelation and sparsification of representations might rely on learning rules that combine Hebbian synaptic plasticity in incoming feedforward synapses with competitive learning of some anti-Hebbian type in the recurrent (inhibitory) connectivity. This may be complemented by some form of intrinsic excitability that acts to avoid unused units (e.g., depolarization of neurons that respond infrequently relative to those that respond frequently). Different approaches to generating such distributed, sparse, and overcomplete neural representations are currently under investigation.

Discussion and Conclusions

We have surveyed and summarized work related to the attractor network paradigm of cortical function, based on simulation results from a large-scale biophysically detailed network model with a modular structure of minicolumns and hypercolumns. We have concluded that such a network model readily performs associative memory and attractor network functions likely to underlie holistic cortical processing. Attractor dynamics may come in the form of spontaneous ongoing activity at about theta frequency, or the network may display a stable ground state that can be replaced by some active coding state triggered by an incoming stimulus. Neuronal adaptation and synaptic depression in the model makes the coding attractor states quasi-stable. The precise sequence of spontaneous recall depends on the overlap structure of the connection matrix in this model. However, in real cortex, additional mechanisms are likely to influence attractor state duration and transitions (e.g., sequence storing connections and loops involving various subcortical structures related to goals, drives, and motivations).

In this context we proposed that working memory is based on fast Hebbian synaptic plasticity rather than persistent activity, which we see as a memory readout process rather than storage per se. Attended and task-relevant objects

are loaded into working memory in a palimpsest fashion such that older memories are written over, and the items thus encoded may be cyclically activated. This view unifies mechanisms behind short- and long-term forms of memory in a biologically plausible manner and makes it quite straightforward to envision interactions between working memory and long-term memory, as well as memory consolidation based on repeated reinstatement.

Fast oscillations in the gamma frequency range emerge readily and influence potentially the dynamics of the network model studied. Spiking statistics during different states show prominent spike synchronization between nearby cells that share the same inhibitory subnetwork; there is also a weak tendency for spike synchrony and phase locking over longer distances. Oscillations and spike synchrony in the gamma range may provide a functional advantage (e.g., promoting low rate firing and quicker and easier transitions between different attractor states). One could expect that a network with continuous STDP learning could induce further fine structure in spiking patterns, like synfire chains. Random networks with STDP learning have been simulated, but this needs to be further examined in a network with attractors stored.

We addressed the important question of whether more than one attractor can be active in the same gamma cycle of about 25 milliseconds, independently and out of phase with each other. We conclude that in our simulation model, lateral inhibition induced rivalry and competition is prominent, thus preventing such uncoupled ongoing activities. However, coordination in the form of theta burst synchronization occurs globally over the entire model network and could be assumed to occur over large parts of real cortex. This might be related to generation of low frequency brain rhythms in the theta–alpha range and phenomena like EEG microstates; it also could be important for dynamic coordination in the form of cross-modal integration, for bottlenecks in perception (e.g., attentional blink), and ultimately for the sequential nature and unity of consciousness.

Future Perspectives

The cortical model of holistic processing discussed here is still incomplete, and work is ongoing to incorporate the missing cortical layers, most importantly layer 4, which presumably provides the circuitry for transforming iconic sensory representations to a sparse and decorrelated format better suited for subsequent attractor network processing. Applying such mechanisms recursively in the cortical feedforward processing stream from one cortical area to the next could be expected to form units in higher-order areas with specific response properties that support some degree of invariant recognition. More large-scale simulations of network models comprising several interacting cortical areas, based on density data and distribution of synaptic connections as well as conduction delays of long-range inter-areal connectivity, are required

to understand dynamic coordination and synchronization phenomena better at this macroscopic scale. Another component function lacking so far in the model, yet crucial for perception and memory, is temporal sequence processing and generation. A plausible addition to the model is Hebbian synapses with a time-delayed and possibly asymmetric pre- and postsynaptic induction window.

We are now beginning to see a constructive partitioning of the main functions of the brain and have had initial success in constructing computational models of its different parts. Phenomena like perceptual grouping, multimodal integration, rivalry, and associative recall from fragmentary information can already be well understood from the perspective of attractor network dynamics, and understanding the interaction between working and long-term memory as well as the mechanisms behind memory consolidation by reinstatement processes seems to be within reach. In addition to perception and memory, there are obviously other equally important major functions to consider: motor control, motor learning, sensorimotor integration, and behavior selection, as well as the emotional and motivational systems so important for goal-directed behavior and learning.

Though it is often convenient to study the different functions separately, efforts must be made to reassemble the modeled results. Integration is necessary to ensure that partitioning is plausible and that the different models can really cooperate efficiently so as to provide sufficient composite functionality. Multiscale modeling at the systems level, using high-performance computers, is necessary to address all of the relevant aspects simultaneously, with a maintained degree of biological detail.

As a final verification, such a whole brain model, or constructed brain, needs to be embodied in the form of an agent and immersed in a complex environment that provides dynamic input and feedback, depending on the output actions generated. The new challenges include, for instance, the detection of task-relevant data as well as the processing involved in the timely selection of the best action, while allowing on-line reinforcement learning to enhance internal representations, motor programs, and stimulus-response mappings to improve performance in future similar situations. Such embodied models will link the computational branch of brain science directly to core problems in information science and technology, including cognitive computing and robotics. An increased mechanistic understanding of brain function will not only have a major impact on improving health in our society; it may also catalyze breakthroughs in information technology, thus paving the way for important applications of novel brain-inspired technology in the service of humankind. Today, this may appear as fantasy and speculation. However, if our knowledge of the brain, together with the capacity and parallelism of computers, continues to develop at the current pace, we may very well reach this stage in a few decades.