IEICE Proceeding Series

Cortical columns for quick brains

Ralph Stoop, Victor Saase, Britta Stoop, Ruedi Stoop

Vol. 1 pp. 852-855
Publication Date: 2014/03/17
Online ISSN: 2188-5079

2012 International Symposium on Nonlinear Theory and its Applications
NOLTA2012, Palma, Majorca, Spain, October 22-26, 2012

Cortical columns for quick brains

Ralph Stoop +, Victor Saase ∗, Britta Stoop ∗∗, and Ruedi Stoop ∗

+ Physics Dept., University of Basel, ∗∗ Inst. of Forensics, Univ. of Berne,
∗ Inst. of Neuroinformatics, UZH/ETHZ, Zurich, Switzerland. Email: ruedi@ini.phys.ethz.ch

Abstract—We use rewiring of neural networks performing real-world cognitive tasks to study whether the particular wiring observed within cortical columns is able to boost neural computation. Upon a vast survey of networks we detect no traces of the proposed effect. It is on the mesoscopic inter-columnar scale that the existence of columns - largely irrespective of their inner organization - enhances the speed of information transfer and minimizes the total wiring length required to bind the distributed columnar computations towards spatio-temporally coherent results.

1. BIOLOGICAL DATA AND BRAIN MODEL

Towards the turn of the 19'th century, J. P. Müller, E. du Bois-Reymond and H. von Helmholtz discovered that neurons are electrically excitable and this predictably affects the electrical state of connected neurons. Shortly after, Golgi and Ramón y Cajal provided their Nobel-prize winning description of neuronal and cortical architecture, revealing in the case of the human neocortex striking columnar structures divided into six layers. Ever since this discovery it has remained a question to what extent neuronal physics and the cortical architecture could account for the exquisite properties of the human brain, at least within the scope limited by Gödel's theorem. We approach this problem by measuring on the inner-columnar scale the effect that the columnar wiring has on real-world pattern recognition tasks in a framework that permits to measure recognition rates without compromising the columnar wiring by the learning process. On the inter-columnar scale, i.e. on the level where the columns interact, the effects on the speed of information transport and computation are analyzed in a framework that offers analytical methods with results valid for sufficiently general situations. Our strategy is to start from networks in which details of cortical architecture are implemented. Using appropriate rewiring, we move away from the biological blueprint, measuring the effects of the removal.

Biological data: Data collected by Roerig et al. [1] for a similar context, evidence in their Log-Log-plot adaption (see Fig. 1) a break between two decay laws at the scale indicated in Fig. 1 by the dashed vertical lines. These lines mark roughly the extension of a (physiological) cortical column. Roerig et al. [1] also noted that 'A small fraction of inputs originated more than one *mm* away'. These observations suggest: A power law or an exponential law of connectivity probability decays with distance across the columnar scale (we chose an exponential decay, which allows the direct comparison of our results to similar work performed in Ref. [2]), a power-law decay for the interaction among whole columns (i.e. on inter-columnar scales) and a much slower decaying power law over very long distances to ensure that the connection probability does not go to zero not too quickly. Motivated by an approximate self-similarity over the microcolumn-column-hypercolumn scales, we will assume that the exponent associated with the slow decay will be close to the one estimated across the columnar distance. Our results do, however, not critically depend on the exactly values of the exponents, only on their relative ordering is of relevance.

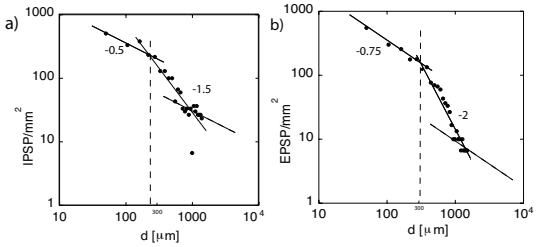

Figure 1: Biological data: Logarithmic density of photostimulation - evoked excitatory (a) and inhibitory (b) synaptic inputs (concentric rings 50 μm apart, from 19 pooled layer 2/3 neurons). Adaptation from Ref. [1]. Vertical dashed lines mark the typical extension of a (physiological) column. Tilted dashed lines: proposed slowly-decaying power-law across large distances

Computation within columns: Despite one hundred years of intense investigations a precise correspondence between functional and physiological columns has not been obtained. Nonetheless, the wide-spread conception is that cortical computation takes essentially place within a column. To test the effect of wiring structure on computation, we therefore will measure first to what extent inner-columnar wiring structure contributes towards computation. We implemented two levels of cortex-inspired wiring (cf. Fig. 2). Upon gradually eliminating architectural details by randomly rewiring the connections, we will measure the impacts that connectivity details have on cognition and computation. Throughout all experiments, the abundance of inhibitory neurons within the population of neurons was kept at 20 percent. To compare with Ref. [3], we will mainly discuss results obtained for networks on a three-dimensional grid of $3 \times 3 \times 15 = 135$ neurons. We

checked, however, that the obtained results also hold for larger networks, and exhibited where this was not the case. In the simple excitatory-inhibitory EI network model, the

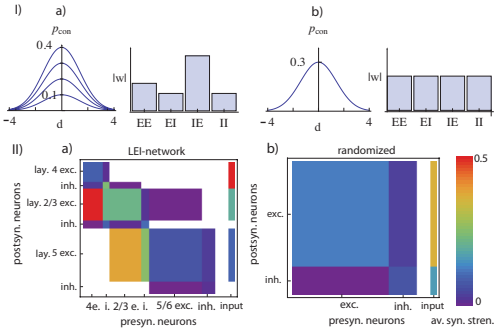

Figure 2: I) a) EI-model, b) EI-control network (uniform synaptic weights w, $\lambda = 2$). p_{con}: probability of a synaptic connection among neurons of distance d for C-values as in the text, w: synaptic strength of the connections. II) a) LEI-model, b) LEI-control network. Input strength to populations is color-coded.

biological architecture is reduced to an excitatory and an inhibitory neuronal population and the connections within and between them. To vary the network structure within the EI model frame, we use a parameter λ ruling the probability for a connection from neuron j to neuron i according to $p_{con}(i,j) = C(i,j) \cdot \exp\left(-d_{i,j}^2/\lambda^2\right)$, where $d_{i,j} = |\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j|$ is the Euclidean distance between the i'th and the j'th neurons' positions in the neural network (see end of paragraph). As λ controls both the number and the typical length of the connections, varying from unconnectedness ($\lambda = 0$) over local next-neighbor connectivity ($\lambda = 1$) to global connectivity ($\lambda = \infty$), this parameter we will use to scan different network structures. $C(i,j)$ establishes the connectivity among excitatory (='E') and inhibitory (='I') neurons, established by means of one pooled synapse. Our choice $C(E,E) = 0.3$, $C(E,I) = 0.4$, $C(I,E) = 0.2$ and $C(I,I) = 0.1$ reflects the typical biological connectivity. Synaptic weights are drawn from a uniform distribution over $[0,1]$, multiplied by the weight factors $w(E,E) = 30$, $w(E,I) = -19$, $w(I,E) = 60$ and $w(I,I) = -19$. The model is compared to a control network where C is uniformly set to 0.3, where the synaptic weights are again drawn from a uniform distribution over $[0,1]$, endowed with a uniform weight factor scaled to match the total weight of the non-control networks and a sign for distinguishing between inhibition and excitation. λ enables us to assess a huge range of neural network architectures. One important feature is that in contrast to classical reservoir computing, we will also vary the fraction I of input-receiving reservoir neurons. In the more detailed LEI network model, also the biological layering structure is implemented. The LEI network is composed from three layers (2/3, 4 and 5/6), each of them containing an excitatory and an inhibitory population. The network is endowed with recurrent connections within the individual layers following the connection probabilities and strengths of Ref. [2]. As in the biological example input mostly feeds into layer 4 (input stream 1 in [2]). Layer 2/3 is the hidden layer, the output neurons are confined to layer 5/6. A family of control networks parametrized by $p \in [0,1]$ is obtained by replacing at each synapse with probability p the pre- and postsynaptic neurons by neurons chosen from the pooled neuronal ensembles of the same kind (excitatory or inhibitory). This rewiring procedure retains the overall connectivity and weight distribution between the excitatory and inhibitory populations, but gradually removes the three-layered structure.

The measurement of the computational effect of inner-columnar wiring structure is done within the framework of a reservoir or liquid state computing neural network ('LSN'). While not in all aspects of top-class efficiency among the possible network types, LSN has successfully been used in robot motion planning. In these networks, learning is confined to the network's periphery. This allows us to assess the pure effect of the inner-columnar wiring structure on computation, without being compromised by the learning process. LSN associates k pairs $\{\mathbf{u}(t)_i\,,\,\mathbf{y}(t)_i\}_{i\in\{1,..,k\}}$ of input / output sequences of individual sequence length T_i (so that $t \in \{1,..,T_i\}$). The dimensionality of the input vectors is denoted by N_u, the dimensionality of the output vectors by N_y. Upon stimulation by the input sequence $\mathbf{u}(t)_i$, the reservoir of N_x neurons generates a state vector $\mathbf{x}(t)_i$ of the same dimension. Let $T = \sum T_i$ denote the total time spanned by the input patterns, and let $\mathbf{X}$ be the $N_x \times T$-matrix of states. Let $\mathbf{Y}$ denote the $N_y \times T$-matrix of the associated patterns. The desired relation $\mathbf{W}_{out}\mathbf{x}(t)_i \simeq \mathbf{y}_d(t)_i$ leads directly to the least-squares optimized read-out matrix $\mathbf{W}_{out} \simeq \mathbf{Y}\mathbf{X}^+$, if $\mathbf{X}^+$ is the (Moore-Penrose) pseudo-inverse of $\mathbf{X}$. During the learning phase, the input and the desired output signal are fed into the reservoir. Care is taken that the scaled input signal optimally stimulates the respective target neuron models (see below). After a transient phase, the optimized output matrix $\mathbf{W}_{out}$ is calculated. This step corresponds to the learning process in classical neural networks. Network realizations are captured in the connection weight matrix $\mathbf{W}$, whereas 'learning' in the traditional sense is confined to the read-out matrix $\mathbf{W}_{out}$. Excitatory connections are reflected by positive synaptic weights $\mathbf{s}$, inhibitory connections by negative weights and absence of connections by zero weights. For all networks, the average neuronal activity is determined by the overall scaling of the inputs and the synaptic weights and by the wiring matrix W. To ensure that networks of a similar level of neuronal activity are compared, the chosen matrices W were scaled to obtain a common largest eigenvalue (1 for EI networks, 0.2 for LEI networks). Matrix W_{in} was chosen by drawing from across $[-0.2, 0.2]$ uniformly distributed random numbers (for Izhikevich neurons scaling by 30 was needed to arrive at the standard parameter scale). The input and synaptic

efficiencies were scaled so that neurons could be excited by their presynaptic partners without reliance on input, but that the firing rates in response to excitation from both input and presynaptic spikes were also sufficiently distant to saturation ($f_{sat} = 1/\tau$). By this we ensured that during all the parameter sweeps, the network was confined to the same dynamic range. In the original LSN, for every input vector an output vector is generated, so that the read-out is memoryless ($'ml'$). For classification tasks it may be advantageous to have a memory span of a size comparable to the stimulus length. Otherwise, LSN will confuse stimuli containing similar parts (e.g. phonemes in speech recognition). This is why we also implemented a more advanced read-out method, where for every neuron and stimulus, an integral stimulus-averaged firing rate provides the read-out vector (integration readout ($'int'$)). To arrive at statements that are largely independent of the network elements, we tested rather distinct models of the neuronal membrane voltage dynamics of which we report exemplarily the results from leaky integrate-and-fire and from fast simple Izhikevich spiking neuron dynamics.

Computation by interacting columns: When zooming out from the columnar dimension to the inter-columnar scale, we have to wrap up the columnar computation and relate it to the computations performed by other columns. Rulkov [4] has demonstrated that any desired neuronal firing behavior representing columnar response can be expressed by a suitably chosen discrete map. The natural model then to use is a coupled map lattice [5] of chaotic maps, to have the response flexibility required for communication. In our modeling of bio-inspired connectivity, the probability that two lattice sites i, j of distance $d_{i,j}$ are connected is $p_{i,j} = \theta \cdot {d_{i,j}}^{-\alpha} + (1-\theta) \cdot {d_{i,j}}^{-\beta}$, which specifies the connectivity matrix, see Fig. 4a. By choosing α, β and θ, a range of network architectures similar to those explored in the LSN paradigm can be accessed. Given $\theta = 1$, the system can be changed from a globally coupled network ($\alpha = 0$) into a nearest-neighbor coupled network ($\alpha \to \infty$). For $0 < \theta < 1, \beta = 0$, $\alpha \to \infty$, the network is coupled to the nearest neighbor with probability 1 and to all other nodes with probability $(1-\theta)$ up to the cutoff M. As a result we obtain a combined nearest neighbor- and random-coupled network. For all intermediate values of α and β, the network is fractally coupled. The cutoff value M determines, together with the underlying topology, the average number of connected nodes k. The interaction of the local chaotic site maps f is modeled by diffusive coupling.

The computation performed at this scale will be characterized by the network's ability to quickly propagate incoming information and to generate among the affected cells a coherent state, expressing that 'computation' has emerged [6]. Both abilities depend crucially on the number of connections k that impinge on a cell. Spatio-temporal information propagation is the maximal velocity of the propagation of perturbations through the network. As a basic approximation, this propagation is the result of two independent contributions: The chaotic instability of the map leads to an average exponential growth of the initial infinitesimal perturbation d_0 applied at site 0, whereas the diffusive coupling results in a Gaussian spreading. The combined perturbative effects at site i are then expressed by the equation [7] $|\delta_{x_i}(t)| \approx d_0/\sqrt{4\pi D t} \cdot \exp(\tilde{\lambda}t - \frac{i^2}{4Dt})$, where D denotes the diffusion coefficient and $\tilde{\lambda}$ is the Lyapunov exponent of the site map. The velocity v of the traveling wave front is determined at the borderline of damped and undamped perturbations, leading to $v = 2\sqrt{\tilde{\lambda}} \cdot \sqrt{D}$. For identical local site maps the task reduces to estimate D from the network topology alone, e.g. via the Markov chain mean transition time. The coherent computational state is expressed by the cells' ability to synchronize in a generalized sense, which requires a minimal number k of connections to impinge on a given site. Full dynamical synchronization of the chaotic sites continues to exist if the condition [8] $| e^{\tilde{\lambda}} - \varepsilon\mu_k | < 1$ (where μ_k are the nonzero Eigenvalues of the graph Laplacian) is maintained. This simple criterion may be overly severe, but can serve as a guideline. Different network architectures should therefore be compared under the constraint of an equal average number of connections k. Biologically relevant indicators of network efficiency will then be the speed of information transfer through the network and the total wiring length required for synchronized columns. Finer network features can be implemented via the connectivity matrix.

2. RESULTS

For the computation within columns, two popular time series classification tasks serve as real-world benchmarks, in contrast with the more abstract computations considered in Ref. [2], Single Arabic Digit speech recognition (13 Mel Frequency Cepstral Coefficients for 10 classes of digits spoken by 88 subjects) and Australian Sign Language ('Auslan') recognition (22 parameters for 95 signs, recorded from a native signer using digital gloves and position tracker equipments). Two general observations emerge from Fig. 3. Whereas the particular neuron models (and the underlying circuit parameters) are of secondary influence (blue vs. red curves), the integration readout (right panels) has a clear advantage over instantaneous readout (left panels). The results for the EI-network demonstrate that the connectivity expressed by λ does not enhance the computational power of the network. Having no recurrent connections among the reservoir neurons does not hamper the recognition rate, suggesting that extremely little computation is owed to synaptic interaction. Low recognition rates from memoryless readout could be from applying the input signal to all neurons, thus constantly overwriting memory that otherwise would be retained in 'hidden' neurons. Therefore we examined in Fig. 3's second row the role of the hidden neurons, by measuring R for networks having a reduced fraction I of input signal receiving neurons. If hidden neurons were beneficial, we should again expect a maximum of R for some optimal value of I. In the Arabic

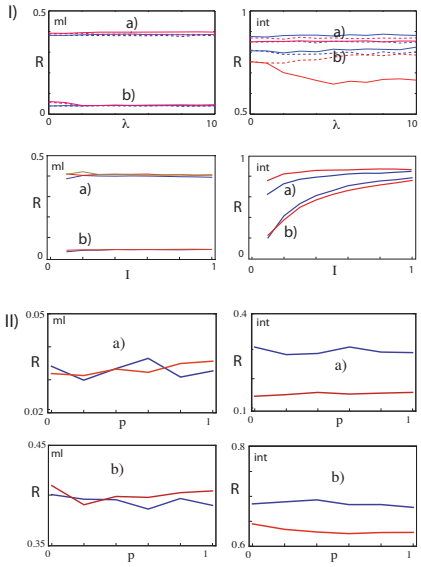

Figure 3: Recognition rate R for a) Arabic Digit, b) Auslan recognition, using leaky integrate and fire (blue curves), or Izhikevich (red curves) neurons, averages over 20 experiments. Left column: memoryless ('ml'), right column: integration ('int') readout. Networks (cf. Fig. 2): I) EI network, dependence upon rewiring via parameter λ (control networks: dashed curves), and on the ratio I of input receiving neurons, at local connectivity (i.e., $\lambda = 2$). Ocher: Izhikevich neurons with $\lambda = 0$ (no connections). II) LEI networks, dependence on the rewiring probability p. $p = 0$: layered, $p = 1$: homogeneous control network.

Digit task with memoryless readout we do not observe a dependence on the number of actually used neurons (i.e. beyond $I = 0.1$, where we have on average 13.5 input receiving neurons, at an input dimensionality of 13). The similarity of the results obtained for $\lambda = 0$ and for $\lambda = 2$ suggests that the nonlinear interaction among the input receiving neurons does not significantly enhance performance. In the Auslan task we see a monotonous dependence of R on I, because for most values of I the number of input receiving neurons is smaller than the input dimensionality (i.e., we have $I \cdot 135 < 95$). The EI network with biology-motivated wiring structure thus does not perform significantly better than the control network. The results obtained for the LEI networks reflecting to more details the columnar layering structures (Fig. 3 II) corroborate the observations made for the simpler model: A significant dependence of R on the rewiring probability p is not observed. These observations are compatible with earlier findings for LSN.

Computation by interacting columns is characterized by the time and the material (i.e., the network's total wiring length) required to obtain a spatio-temporally coherent effect at some distant information-processing network site. We measured the speed of information transport through the network for doubly fractal, single fractal, random, and nearest-neighbor, coupling topologies, and found that the doubly fractal architecture suggested by the data of Roerig

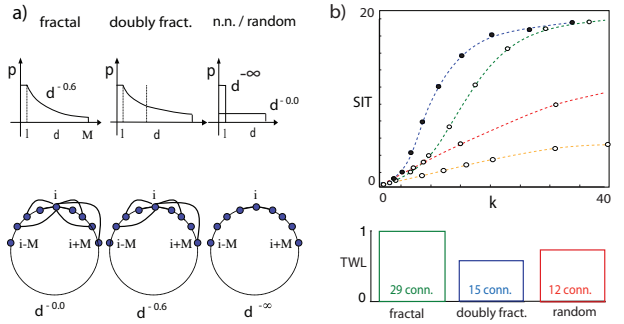

Figure 4: a) Main network connectivities compared (p: connection probabilities, d: distance, M: cutoff, see text). Below: examples of 'fractal' wirings. b) 'SIT' as a function of average number of connections k. From top: doubly fractal ($\theta = 0.2$, $\alpha = 0.5$, $\beta = 2.0$), fractal ($\theta = 1$, $\alpha = 0.7$), random, next-neighbor topology. Networks sizes: $N = 4096$. Averages over 100 experiments. Lower panel: Typical number of connections k required for synchronization and (histogram height) relative total wiring length TWL. $N = 512$, 10 realizations.

et al. consistently enhances speed of information transfer if compared to the alternative networks (Fig. 4b), upper panel). In this figure, we plot the speed of information transfer $v =:$SIT in arbitrary units. Speed enhancement persists across a wide selection of pairs of exponents as long as the qualitative size of the exponents is preserved. Our numerical experiments show that under the condition of synchronization, the doubly fractal architecture jointly optimizes total wiring length TWL and speed (Fig. 4b), lower panel).

References

[1] Roerig B, Chen B (2002) Relationship of Local Inhibitory and Excitatory Circuits to Orientation Preference Maps in Ferret Visual Cortex. *Cerebral Cortex*, vol. 12, pp. 187-198.

[2] Haeusler S, Maass W (2007) A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models. *Cerebral Cortex*, vol. 17, pp. 149-162.

[3] Maass W, Natschlaeger T, Markram H (2002) Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Computation*, vol. 14, pp. 2531-2560.

[4] Rulkov NF (2002) Modeling of Spiking-Bursting Neural Behavior Using Two-Dimensional Map. *Physical Review E*, vol. 65, 041922.

[5] Bunimovich LA, Sinai YaG (1988) Spacetime chaos in coupled map lattices. *Nonlinearity*, vol. 1, pp. 491-516.

[6] Stoop R, Stoop N (2004) Natural computation measured as reduction of complexity. *Chaos*, vol. 14, pp.675-679.

[7] Cencini M, Torcini A (2001) Linear and nonlinear information flow in spatially extended systems. *Physical Review E*, vol. 63, pp. 056201-1-056201-13, DOI: 10.1103.

[8] Atay FM, Jost J, Wende A (2004) Delays, Connection Topology, and Synchronization of Coupled Chaotic Maps. *Physical Review Letters*, vol. 92, pp. 144101.