Toward tractable AGI: Challenges for System Identification in Neural Circuitry

Randal A. Koene

Carboncopies.org
1087 Mission St., San Francisco, California 94103, USA
tel: +1-650-388-0725
Randal.A.Koene@carboncopies.org

Abstract. Feasible and practical routes to Artificial General Intelligence involve short-cuts tailored to environments and challenges. A prime example of a system with built-in short-cuts is the human brain. Deriving from the brain the functioning system that implements intelligence and generality at the level of neurophysiology is interesting for many reasons, but also poses a set of specific challenges. Representations and models demand that we pick a constrained set of signals and behaviors of interest. The systematic and iterative process of model building involves what is known as System Identification, which is made feasible by decomposing the overall problem into a collection of smaller System Identification problems. There is a roadmap to tackle that includes structural scanning (a way to obtain the "connectome") as well as new tools for functional recording. We examine the scale of the endeavor, and the many challenges that remain, as we consider specific approaches to System Identification in neural circuitry.

Keywords: system identification; whole brain emulation; functions of mind; measurement tools; neurophysiology

1 Tractable AGI through System Identification in Neural Circuitry

Artificial General Intelligence (AGI) is, at a minimum, a system that is able to deal with challenges or tasks arising in circumstances of our natural environment. It is possible that there are elegant mathematical approaches to AGI that

address those minimum requirements and are theoretically sound [1]. Theoretical soundness does not imply practical feasibility. The most elegant mathematical methods can be the most compute-hungry, slow, impractical solutions. From a practical standpoint, there is much to be said for short-cuts that are suitable to the environment and the challenges.

One system that contains many of those short-cuts and that we often think of in terms of AGI is that of the (human) mind. Deriving from a brain a functioning system that implements a degree of intelligence and a degree of generality within the constraints of compatible environments may be done at several different levels of cognitive abstraction. The level that most interests me and many colleagues is that of computational neurophysiology, systems of neuronal circuitry [2].

In part, this choice comes from the fact that neuroscience has spent the past 100 years gaining experience at that level and devising functional representations that are well grounded in identified physiological mechanisms. The other reason for this choice is that our goals are in some ways a reversal of the quest for practicable AGI. We begin with a system that has many specific short-cuts built in that give it satisfactory performance under current real-world circumstances. But our interests involve making that system more adaptable to novel environments and challenges [3].

In this paper, we highlight the importance of good System Identification [4]. We point out what choices need to be made and which tools may be applied. Most importantly, we identify the significant challenges that appear throughout the process of System Identification and due to the need to integrate efforts with several different types of tools.

2 Representations and Models

The exact sciences depend on improving understanding by describing observed effects through representations and models. Some things about nature are predictable. Pieces of nature exist within an environment. There, the various pieces are not wholly independent. Conditions of some piece at some time predict aspects of the conditions in another piece. We say they affect each other. There are signals between the pieces that convey information. We want to understand more about the predictable dependencies, so we explicitly describe the signals and how the information they convey is processed.

Behavior and Signals of Interest

Nature has an awful lot of pieces and descriptions become quite complicated. Systematic and iterative improvement of a description is model building. Initially, we keep it simple, we constrain our models. There is an effect of particularly interest. Ideally, we focus solely on the scope and details that are needed to explain that effect. In neuroscience, the interesting effects are often called (task-specific) behaviors. E.g., object recognition, emotional responses, executive decision making, and even conscious or aware behavior. In AGI, there are

also particular effects or behaviors that are interesting and for which we want to carry out System Identification in the brain.

Now we know our piece of interest, e.g., a molecule of gas or a neural circuit in the brain. We look at how that piece may be communicating with others. What are the signals that could be involved in the effects? Overall, physics describes all interactions in terms of four types: gravity, electromagnetism, weak nuclear force and strong nuclear force. While those are a limited set, we can constrain their manifestations further and consider electric current, electromagnetic radiation, etc. A piece of neural tissue may respond to (ionic) electric currents, temperature (gradients), pressure or shearing forces, sonic transmissions, electromagnetic fields, and more. Experimental work helps us to create a priority order. By and large, most signals appear to drown in noise, losing predictive value. Electric currents, and in particular the powerful discharges known as neural action potentials or spikes appear to carry the dominant information [5].

Discovering the Transfer Function

In Control Theory, the piece of nature being modeled is sometimes called a black box, which has state, receives input and produces output. The process of updating its state and generating output is described mathematically by a transfer function. When we find suitable transfer functions we learn about the black box in the context specific behavior and signals. There are numerous formal methods, and a general example of one that has been successfully applied (e.g., in Ted Berger's neuroprosthesis [6]) is to find the kernels of a system that is expressed as a discretized Volterra series expansion, as in Eq. 1. The kernels, H_n, express the contributions of a history of input, $\mathbf{x}$, to system output $f(\mathbf{x})$, with a finite number of mn coefficients $h^{(n)}_{i_1 \ldots i_n}$.

$$
\begin{aligned}
f(\mathbf{x}) &= H_0\mathbf{x} + H_1\mathbf{x} + H_2\mathbf{x} + \ldots + H_n\mathbf{x} + \ldots + \mathbf{H_m}\mathbf{x}, \\
H_n\mathbf{x} &= \sum_{i_1=1}^{m} \ldots \sum_{i_n=1}^{m} h^{(n)}_{i_1 \ldots i_n} x_{i_1} \ldots x_{i_n}.
\end{aligned} \tag{1}
$$

3 Mental Processes and Neural Circuitry: Brain Emulation

The effects that interest us are those that we associate with our experiences: Sensory Perception, Learning and Memory, Problem Solving and Goal-Directed Decision Making, Emotional Responses, Consciousness and Self-Awareness, Language Comprehension and Production, Motor Control. Some are externally observable and some are part of the internal experience. Neurophysiologically, these involve the interactions of ensembles of neurons within a specific circuit layout.

System Identification in Neural Circuitry

There is no consensus about exactly which signals are or are not essential to brain function, but we take an iterative and systematic approach. We make initial assumptions about signal to noise ratios, about the sort of output that reliably affects the environment during interesting behavior, and about the sort of signals that neurons are well-suited to deal with. Biophysical mechanisms of sensory input (e.g., at the cochlea, at the retina) produce electric nerve signals characterized by trains of fairly uniform neural spikes with very specific rates and time intervals. Similarly, the primary output through muscle control (e.g., vocal cords) employs trains of neural spikes. Finally, a primary means of long-term state-change (ie., learning) is governed by modified synaptic strength. That modification also depends crucially on the temporal order and time-separation between pre- and postsynaptic neural spikes [7]. A representation that successfully predicts spike times may therefore be a good first iteration of a model of system processes in neural circuitry.

One result that was achieved with these assumptions is demonstrated in the cognitive neural prosthesis devised by the lab of Theodore Berger (UCS). Using System Identification in a Volterra series expansion, they developed a chip that contains a multi-input multi-output model with non-linear parameters that are specified after learning from consecutive presentations of spike data. The input of the system is obtained through an array of electrodes in region CA3 of the hippocampus, while the output is delivered to region CA1 [8]. These regions are crucial in the formation of new declarative and episodic memories. The chip is designed to alleviate dysfunction caused by stroke, trauma or disease.

A more general technique designed to work with out initial assumptions was developed by Aurel A. Lazar and Yevgeniy B. Slutskiy and is called the development of Channel Identification Machines [9]. It is a formal method to identify a channel – modeled as a multi-dimensional filter – in a system where a communication channel is cascaded with an asynchronous sampler. The samplers consist of neuroscience or communication models, e.g., integrate-and-fire neurons, asynchronous sigma/delta modulators, general oscillators with zero-crossing detectors. A channel can be approximated to an arbitrary degree of precision and the method was generalized and applied in noisy conditions.

4 Simplification of an Intractable System into Collections of System Identification Problems

Meaningful System Identification that could reproduce both observable behavior and internal experiential states of an entire brain is entirely unfeasible when the complete system is treated as the black box. This has to be broken down into many black boxes that communicate with one-another. We need: a.) to choose smaller black boxes, b.) to acquire enough data about I/O correlations at those smaller black boxes for their System Identification, and c.) to know the relevant communication that is possible between those black boxes.

We can address c.) by looking inside the system, noting locations of the smaller components and tracing the connectivity between them. For spike trains, the communication pathways are dendrites and axons, and the synapses where they meet. The new field of Connectomics in neuroscience deals with this problem [10]. For other effects, such as extracellular field potentials and diffuse neurotransmitters, the surrounding medium and emission and diffusion may be taken into account.

A well-known choice for a.) that contains very tractable sub-systems is decomposition of neurons into the electrical compartment analogs of a so-called compartmental neural model (Fig. 1). The I/O data that can be obtained largely determines if this, or a another level such as whole neurons, is the appropriate level of simplification. High-resolution connectomics by electron microscopy obtains the morphological data for compartmental modeling. There are a number of labs working on this and in 2011 the approach resulted in proofs of concept by Briggman *et al.* [11] and Bock *et al.* [12], using data from the lab of Winfried Denk (Max Planck).

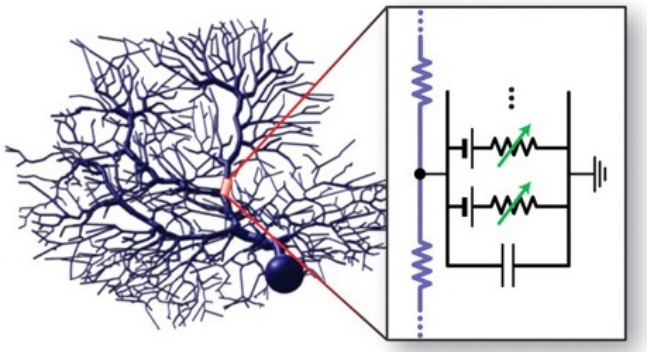

Fig. 1. The compartmental model of a Purkinje Cell. The electric cable analogy for one compartment is highlighted in the box.

To satisfy b.) and properly characterize the response of a neuron we need observations that allow us to set and test parameter values, which relate to the sensitivity and manner in which input currents affect neuronal membrane potential, the resulting action potential once a threshold potential is reached, and the time-course of restorative dynamics (e.g. after-hyperpolarization, after-depolarization).

Tools for Structural Decomposition

In neural tissue, sensible boundaries must be drawn around pieces of the neural circuitry, and I/O contacts between the pieces must be identified. A geometric decomposition into 3D stacks of voxels of equal size is one approach, such

as through magnetic resonance imaging (MRI). Another method is to identify neural cell bodies, as in slice or culture on top of an array of electrodes, and to use the correlations between measured activity at each cell body to derive a functional connectivity map.

Anthony Zador is developing a biological protocol to derive the target neurons of any neuron. Zador uses biological markers such as unique sequences of RNA or DNA to mark the pre- and postsynaptic sites of synapses. The markers act as bi-directional pointers [13]. But the most detailed and successful tools to date section or ablate pieces of brain tissue and take electronmicrographs at resolutions up to 5nm from which 3D geometric morphology can be reconstructed. Even individual synapses can be identified. Excellent results have come out of the labs of Winfried Denk (Max Planck), Jeff Lichtman (Harvard) and Ken Hayworth (Janelia Farms).

Data from Structure

System Identification for individual neural compartments can use standard models that employ an electric cable analogy and Hodgkin-Huxley equations. Morphology can provide some insight into functional behavior. For example, compartment radius and length constrains the conductance of electrical currents. Morphology can also categorize a neuron or synapse, which constrains the possible response functions. Despite these constraints, even small systems contain numerous parameters. Not all of those relate directly to visible and unique morphological features. Even where they do, the reliability and precision of measurements may not be adequate.

Parameter Tuning Among Connected Systems: Reference Points

Parameters must be tuned such that sub-systems behave sensibly on their own in in cohesion with connected neighbors. We can do System Identification for signals of interest at a black-box by observing activity, or at a gray-box when we can stimulate and observe. Tuning and verification involves measurements at reference points.

If the resolution of reference points is less than the resolution of structural decomposition then System Identification depends on our ability to map measurements to a collection of sub-systems and the combinatorial size of the collective problem. In how many ways might the sub-systems be interacting to produce observed responses? We may not be able to determine system parameters uniquely if that number is large. The amount of observations needed and the duration of observation increase with complexity. Clearly, there is great value in having tools that provide measurements at many more reference points, ideally at a resolution that approaching the resolution of the sub-systems.

Tools for Characteristic Reference Recordings

There is now strong interest among neuroscientists in the development of tools for high-resolution in-vivo recording. Arrays of thousands of recording electrodes

are being developed and combined with optogenetic techniques so that selective observation of specific groups of neurons can be guaranteed. Microscopic wireless probes and functionalized nanoparticles with simplified task-specific capabilities are being developed to counter some of the disadvantages of extensive tethered electrodes. There is also a collaborative effort underway to create biological tools that employ DNA amplification as a means to write events onto a molecular "ticker-tape" [14]. The project goal is to be able to record signals from all neurons in a brain, and potentially to measure at resolutions beyond that.

5 Challenges

Some challenges are general to System Identification. Some are particular to neurons and neuronal models. There are unique challenges that arise when working with pieces of neural tissue and consequent large neural circuit models. And some challenges are exclusive to the domain of whole brain circuit reconstruction. Many of those involve the integration of techniques for data acquisition from structure and function that are developed with the constraints of particular novel tools.

Signals and Predicting Spikes

A careful assessment of the System Identification problem for the experiences we wish to represent demands that we consider contributions outside the domain of neural spiking. For example, are the experiences meaningfully represented by states of cells other than neurons, glia perhaps? Or, are there significant ways in which neurons influence each other even in the absence of spiking [15] – can neurons relate to each other without receiving spikes or activity directly caused by spikes? Our initial assumption is that predicting spiking within an acceptable error range implies good emulation (Fig. 2). Spikes are not epiphenomenal, but rather the currency upon which the rest of sensation rests. Spikes precede ensemble responses and field emissions. An important challenge is to test these assumptions.

Good temporal spike predictors demand that we observe or deduce when spikes would occur in the original system. Additional information, such as membrane potentials and influences on such can give improve our ability to build good local predictors.

Low-res Validation, 3D Reconstruction at 5nm and Plasticity

The snapshots of baseline or differential activity in large volumes of tissue that are provided by MRI are too imprecise for parameter tuning, but they can be a means of model validation. The model should produce a sensible virtual MRI in terms of distribution and propagation of activity. A challenge is that aligning a virtual MRI generated by the model with actual data demands that the model also replicates the expected 3D spatial geometry.

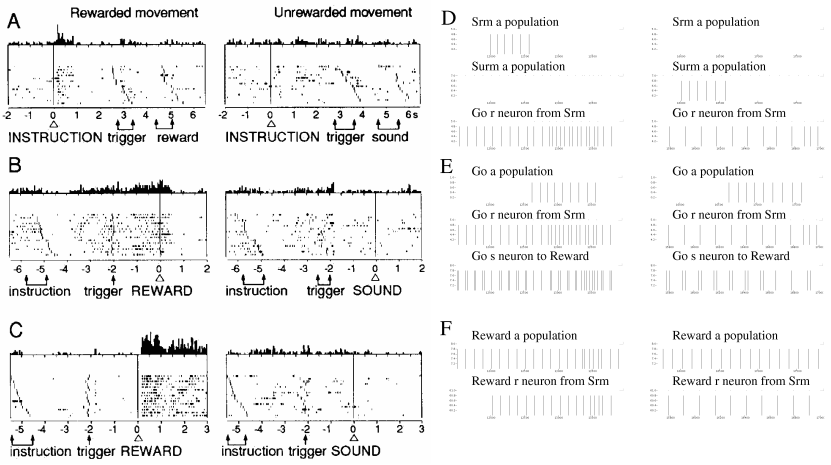

Fig. 2. Spike prediction is functional emulation [16].

Detailed geometric and morphological data is provided by 3D reconstructions from EM scans at resolutions up to 5nm. We can tell if a cell is a pyramidal cell or an interneuron, which helps model activity dynamics and the receptor channels that are likely present. Still, component identification is challenging, because classification presently relies entirely on morphology. There is an effort to add a direct means of protein identification, which would alleviate this problem.

When reconstructed in terms of compartments, radius and length of a cylinder gives estimates of resistance and capacitance, although those estimates also depend on the model of the identified type of a neuron. Measurements are subject to a degree of reliability. Averaging is possible, but that does not remove cumulative systematic errors. Some measurement data is likely to be entirely unrecoverable.

Brains are plastic. Mostly, we think of plasticity in terms of learning [17], modifications of synapses and even of the available connections (Fig. 3). But there is also plasticity in terms of deformation. Apparently, some aspects of morphology are relevant to model building, while others are mere features of the snapshot taken during scanning and reconstruction. Present tools for structural connectomics offer little insight into the temporal dynamics of these gradual changes in neural circuitry.

Ticker-Tape Data and Interference during Measurement

Tools that functionally characterize activity at reference points should give some insight into temporal dynamics and memory. The ticker-tape approach may even be able to record from all neurons simultaneously. Encoding by means of voltage-dependent increases in the error-rate of DNA amplification is not entirely reliable, which is compensated by using multiple tapes per cell. Recordings may be

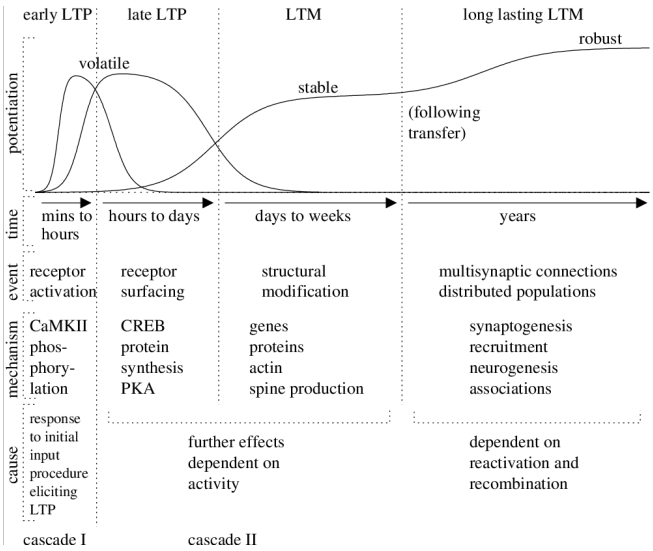

Fig. 3. There is a cascade of different memory mechanisms that implement brain plasticity for learning. Representations of these should be included in a model, but most of those are invisible to structural snapshots.

synchronized by time-stamp signals, and can identify spike times and possible even voltage levels. But the method of data recovery poses a challenge when combined with tools from other projects in the endeavor. DNA snippets are extracted from cell bodies and the process does not retain tissue samples that could be scanned structurally by electron microscopy. How do we obtain the structural connectome, and how do we know which part of the ultrastructure a molecular tape came from? Scanning of slices prior to DNA extraction might be possible if special care is taken in the method of fixation of brain tissue. There is also some question whether the presence of many molecular ticker-tapes might interfere with cell function.

Interference challenges also appear when functional data is obtained by fluorescent microscopy with calcium dyes or voltage sensitive proteins. Using those tools to obtain full coverage throughout the neural tissue and in complete brains involves significant disturbance in the form of view ports and insertion of microscope devices. This problem is similar to the one face by large electrode arrays.

Microscopic Wireless and Data Quantities

Microscopic wireless electrodes and functionalized nanoparticles are feasible alternatives where each individual probe has strict task constraints. Challenges are the possible power requirements and demands of data delivery. These may make it difficult for a whole network of probes to measure continuously at a rate that captures all interesting events. Functional characterization by these methods is

simplified when done by sporadic sampling from different locations until each location is adequately characterized. When has enough data been collected and how are the results validated? Functional probes may help us look at temporal dynamics, but it can be difficult to ensure their location within the tissue over extended time spans. Frequent spatial registration is likely necessary. Ultimately, a microscopic functional probe technique can be combined with structural data acquisition by electron microscopy, because the probes can remain in the tissue when it is scanned.

When we have a way to record spikes, electric field potentials or membrane potentials from all neurons in a piece of neural tissue we still need to know what is a sufficient sample set. And what is the required sample rate? Can we predict neural dynamics from the observation of the shape of an action potential response? Should we observe the responses to a collection of possible input combinations in order to estimate connection strengths and predict spike times? Do we need to run the cell through its paces with a full battery of stimulation protocols? Does tractable System Identification demand that we do so at a higher resolution, on pieces of dendrite? Can we map lower resolution activity data to high resolution structure data so that compartment parameter values are sufficiently constrained?

Virtual and Small System Proof of Concept

Many of the challenges listed above can be dealt with confidently only if the process of System Identification is tested by iteratively building incrementally improved models of small systems. A small system can be a proof of concept that demonstrates steps of the process and overall feasibility. There are some small systems that are receiving attention at this time: Several groups are working on the nematode C. elegans (e.g. D. Dalrymple). Others are reconstructing pieces of retina (e.g. Briggman *et al*). Neuroprosthetic applications are being built for pieces of the hippocampus (T. Berger) and for the cerebellum (S. Bamford). There is also a project to extract memory directly from a piece of neural tissue (S. Seung)

Sometimes, we can also carry out virtual process testing. Programs such as NETMORPH [18] are able to "grow" or generate virtual neural tissue, with a known structure (Fig. 4) and known characteristic functions. We can explicitly test algorithms used to set parameter constraints from structure data, and we can test algorithms that take partial functional data and tune parameters accordingly. The results can give an indication of the minimal functional data that needs to be collected, and they can point out limitations in reconstruction from morphology. More abstract calculations of boundary conditions may also be possible, deducing constraints set by structure and additional information provided by patterns of input and correlated output.

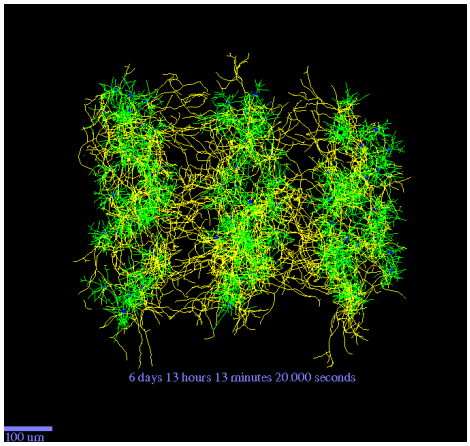

Fig. 4. Structurally detailed network generated with NETMORPH.

6 Conclusion

It is often impossible to properly gauge which difficulties will turn out to be significant problems unless you work your way through the entire process. That is a main reason why proof-of-concept systems are so important.

System Identification is not a new field. It is done in every area of the exact sciences and engineering. Undoubtedly, most of the problems encountered when working with neural tissue are not entirely novel either. Examples of similar problems and the solutions that are employed may be found in other fields.

From the discourse above, it should be clear that while it is important and useful to build tools that acquire high resolution structure data and that acquire high resolution spatial and temporal functional data, that is not the whole solution. Other significant challenges are the integration from different data sources, turning a sea of data into parameter values, and validating those values.

A goal of this paper was to describe what System Identification entails in the case of reconstructing brain circuitry, and to communicate the reality of this effort beyond the confines of the discipline. Hopefully, this will lead to input from many other experts in the area of System Identification, which will lead to a better understanding of the problems and an improved roadmap to solutions.

References

1. Hutter, M.: Universal algorithmic intelligence: A mathematical top-down approach. In Goertzel, B., Pennachin, C., eds.: Artificial General Intelligence. Springer (2007) 227–290 2
2. Koene, R.: Fundamentals of whole brain emulation: State, transition and update representations. International Journal of Machine Consciousness **4**(1) (2012) 2

3. Koene, R.: A window of opportunity. In: H+ Magazine. (May 2012) http://www.carboncopies.org/a–window–of–opportunity Based on the TEDx-Tallinn 2012 talk. 2
4. Ljung, L.: Perspectives on system identification. In: In Plenary talk at the proceedings of the 17th IFAC World Congress, Seoul, South Korea. (2008) 2
5. Rieke, F., Warland, D., de Ruyter van Steveninck, R., Bialek, W.: Spikes: Exploring the Neural Code. MIT Press, Cambridge, Massachusetts (1997) 3
6. Hampson, R., Gerhardt, G., Marmarelis, V., Song, D., Opris, I., Santos, L., Berger, T., Deadwyler, S.: Facilitation and restoration of cognitive function in primate prefrontal cortex by a neuroprosthesis that utilizes minicolumn-specific neural firing. Journal of Neural Engineering **9** (2012) doi:10.1088/1741–2560/9/5/056012 3
7. Bi, G., Poo, M.: Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. Journal of Neuroscience **18(24)** (1998) 10464–10472 4
8. Berger, T., Ahuja, A., Courellis, S., Deadwyler, S., Erinjippurath, G., Gerhardt, G., Gholmieh, G., Granacki, J., Hampson, R., Hsaio, M., Lacoss, J., Marmarelis, V., Nasiatka, P., Srinivasan, V., D., S., Tanguay, A., Wills, J.: Restoring lost cognitive function. IEEE Engineering in Medicine and Biology **24**(5) (2005) 30–44 4
9. Lazar, A., Slutskiy, Y.: Channel identification machines. Computational Intelligence and Neuroscience (2012) (In press.) 4
10. Seung, S.: CONNECTOME: How the Brain's Wiring Makes Us Who We Are. Houghton Mifflin Harcourt (2012) 5
11. Briggman, K., Helmstaedter, M., Denk, W.: Wiring specificity in the direction-selectivity circuit of the retina. Nature **471** (2011) 183–188 5
12. Bock, D., Lee, W.C.A., Kerlin, A., Andermann, M., Hood, G., Wetzel, A., Yurgenson, S., Soucy, E., Kim, H., Reid, R.: Network anatomy and *in vivo* physiology of visual cortical neurons. Nature **471** (2011) 177–182 5
13. Zador, A.: Sequencing the connectome: A fundamentally new way of determining the brains wiring diagram. Technical report, Project Proposal, Paul G. Allen Foundation Awards Grants (2011) 6
14. Kording, K.: Of toasters and molecular ticker tapes. PLoS Computational Biology **7**(12) (December 29 2011) e1002291 doi:10.1371/journal.cpbi.1002291. 7
15. Anastassiou, C.A., Perin, R., Markram, H., Koch, C.: Ephaptic coupling of cortical neurons. Nature Neuroscience **14**(2) (2012) 217 7
16. Koene, R., Hasselmo, M.: An integrate and fire model of prefrontal cortex neuronal activity during performance of goal-directed decision making. Cerebral Cortex **15**(12) (2005) 1964–1981 Advanced Access published on April 27, 2005. 8
17. Koene, R.: Functional requirements determine relevant ingredients to model for on-line acquisition of context dependent memory. PhD thesis, Department of Psychology, McGill University, Montreal, Canada (2001) 8
18. Koene, R., Tijms, B., van Hees, P., Postma, F., de Ridder, A., Ramakers, G., van Pelt, J., van Ooyen, A.: NETMORPH: A framework for the stochastic generation of large scale neuronal networks with realistic neuron morphologies. Neuroinformatics **7**(3) (2009) 195–210, doi: 10.1007/s12021–009–9052–3 Published online: 12 August 2009. 10