Journal of Artificial General Intelligence 4(3) 1-9, 2013
DOI: 10.2478/jagi-2013-0012

Submitted 2013-12-31
Accepted 2013-12-31

Editorial: Whole Brain Emulation seeks to Implement a Mind and its General Intelligence through System Identification

Randal Koene RANDAL.A.KOENE@CARBONCOPIES.ORG
Carboncopies.org
1087 Mission Street
San Francisco, CA 94103, USA

Diana Deca DIANA.DECA@LRZ.TUM.DE
Institute of Neuroscience
Biedersteiner Str. 29
D-80802 Mnchen, Germany

1. An Introduction to Whole Brain Emulation

Whole brain emulation (WBE) is a systematic approach to large-scale neuroprostheses with the intent to replicate the functions of a specific mind in some other operating substrate. The engineering practice of system identification can be applied in a way that makes this big problem a feasible collection of connected smaller system identification problems to solve.

Whole brain emulation is an essential goal for neuroscience. Following Richard Feynman's famous 1988 Caltech chalkboard quote: "What I cannot create, I do not understand." To create or build a human mind we need models, a combination of building blocks with processes. When we explain something that is observed, e.g., mental functions and behaviors, we strive to make that predictable within constraints that satisfy our interests: We create boundaries, we measure within those well-defined outlines, and then we use those measurements to derive model processes enabling outcome prediction. Within the defined system outlines of our model, taking into account defined sets of signals, we mathematically describe interactions (which may be expressed in information theoretic terms).

Every aspect of modern science relies on creating representations of things. In each case, we focus on the signals and the observables (or behavior) that interest us. Then, we try to interpret in terms of functions what the system processes are doing. Where brain functions are concerned, some cognitive prosthetic work, such as the pioneering efforts of the labs of Theodore W. Berger at the University of Southern California, has managed to carry out these steps and produced successful experimental results (Berger et al., 2012). Berger's team has developed and tested an experimental hippocampal neural prosthetic that is implemented on a bio-mimetic chip. A transfer function was identified and used to replicate the operational properties of biological neural circuitry in a region of the rat hippocampus known as CA3. In experiments, the prosthesis is able to reproduce the way in which input to the region is turned into output from that region. This method of developing neuroprostheses, with demonstrated success in rats, is presently being tested in primates (Marmarelis et al., 2013).

1.1 System Identification

Brain emulation strives to achieve a functional re-implementation by which it is possible to predict an active brain state and behavior at a time $t + dt$ (with acceptable error) if we know the state at a slightly earlier time t. This process of discovering the functions by which an unknown system, turns input into output is often called system identification (Ljung, 2008): Investigating the correlated input and output, then attempt to determine which functions constitute characteristic processing. Of the unknown system that is the brain, we know that it is composed of many physiologically similar components, such as neurons of several types and synapses of several types. We also know that it contains a very large collection of such components and that their arrangement is highly complex.

To carry out system identification, we need to observe a working system during its exposure to a sufficiently complete series of input patterns. We can then describe transfer functions and expected output with the inclusion of all relevant system behavior. When a system contains more internal state, receives input through more channels, and produces output through more channels, we have to make many more observations. If an entire mammalian brain is approached as a single unknown system then we would probably have to observe its input and output throughout its entire life-span. Even then, what we could deduce from the resulting data would be flawed and would fail to capture much latent function. Instead, the whole brain problem needs to be broken down into smaller pieces, into constituent sub-systems that communicate with one-another. Ideally, the result is a collection of individually manageable system identification problems that are a good fit the tools at our disposal with which we measure and collect data, build functional models, estimate model parameters and ultimately devise prosthetic replacements.

The strategy involves three steps: 1.) Choose the smaller sub-systems. 2.) Find out how they are connected and communicating. And, 3.) make measurements at each sub-system and identify its system functions. Considering the problems and possible solutions for those three steps of the system identification strategy allows us to work on a roadmap toward emulating functions of brain tissue.

1.2 Four Pillars of Development

Iteratively, we can determine that "sweet spot" where a our ability to solve a collection of connected and individually tractable system identification problems meets our ability to build new tools for high-resolution measurements. At that point, brain emulation at the scale of a human brain is a feasible project. We can categorize areas within a roadmap toward whole brain emulation according to four main pillars:

1. Hypothesis testing – iteratively evaluating proofs-of-concept on our way to the sweet spot;
2. Structure – the decomposition of the system identification problem into many smaller problems, largely by gathering so-called "connectome" data;
3. Function – characterizing each system, an area with tool-development needs that are addressed, for example, in the BRAIN Initiative (Obama, 2013);
4. Emulation – the mathematical representations and computational platforms needed.

As the list shows, a development roadmap includes structural scanning (connectomics) as well as new tools for functional recording that greatly improve upon tools typically available

to neuroscientists today (Deca, 2012). Those investigative technology requirements are being addressed in a number of ongoing projects, several of which we point out in following paragraphs.

We cannot know *a-priori* what are all the relevant contributors to the system processes that interest us, nor can we describe in detail how to make accurate system predictions. However, some units performing the input-output computation have been proposed. In both philosophy of mind and in the history of neurophysiology, major brain areas have been regarded as generating inputs and receiving outputs (e.g thalamic inputs going into primary sensory areas). Another type of unit in computational neuroscience is a network of neurons. It could be that there are networks dealing, at least for a limited period of time, with a specific function (the rabies virus is currently used to check if this is true). Typically, the single neuron has been considered the most fundamental computational unit in the brain, and a lot of experiments have been performed in this paradigm (based on the assumption that a neuron computes sensory-stimulation-related inputs coming from other neurons in the network into one single output). However, a single dendrite may act as a computational unit and generate regenerative events. Single spines and ion channels also possess computational properties (ion channels can act as coincidence detectors for other channels and modify their activity according to specific cues, and spines can block or allow an input to pass through the entire dendrite). All these units are in fact the same unit at different scales. Apart from the terminology, it is important to keep in mind that computation is performed in the brain at different scales and that the causality of this computation can be observed experimentally and understood as the body of evidence grows.

We can begin a formal description of mental processes, while taking care not to be overly restrictive about the underlying mechanisms to be considered. After-all, coming up with a satisfactory model or theory is an iterative process that is based both on conceptualization and on data evaluation. We can make some initial assumptions based on the presence or lack of certain evidential data at this time. For example, we might assume that the brain relies on particular biophysical mechanisms, which should be modeled to adequately predict and replicate the processes of the mind.

Once we have a model of processes that act on signals, and once we acknowledge that boundaries are drawn around sub-systems in some way, then, when we focus on a specific set of sub-systems, we can identify the boundaries between them. For example, at a large scale this could be the boundary between the mental experiences of a person and the environment that is stimulating those experiences (the body and surrounding world that comprise that environment). We can also identify boundaries drawn sensibly at smaller scales, for example, experiences attended to versus undesired/randomized/other experiences, spatially constrained sub-systems such as neurons, temporal discretization (e.g. next-spike prediction), and so forth.

Given such boundaries and signals we can talk about an exchange of input and output. If the sub-systems have been chosen at an adequate resolution to suit our experiential level of description (chosen representation) then a process model can be a transfer function describing the conversion of input to output within each sub-system. The description can include hysteresis (memory) in the sub-system. As we learn how to interpret the conversion of input into output, our description of that system process becomes our understanding of the system Koene (2012a).

1.3 Proof-of-Concept

The concrete success of the proposed systematic approach to brain emulation, embodied by a successful neuroprosthesis, is evaluated with regard to experimental goals and well-define

performance requirements, which can be expressed as experiential criteria (Koene, 2012a). One example of such success is the experimental performance of the hippocampal prosthetic chip, as tested in laboratory settings by Berger's team. Another is the proof-of-concept verification carried out in published work by Briggman, Helmstaedter, and Denk (2011), where the connectome of a sample of retinal tissue was studied. They used the structural data obtained by electron microscopy in the lab of Winfried Denk to derive and predict the functional operations (such as direction selectivity) of specific retinal ganglion cells. The experimental protocol used there resembled a proposed approach for the derivation of brain emulation functions from morphological measurements in neuronal tissue. The publication was an important proof-of-concept, because functional derivations were verified by comparison with functional data that had been gathered in the same specimen through fluorescent optical microscopy.

1.4 Technology

Whole brain emulation relies on determining precisely which signals we care about and then breaking the problem down into a collection of smaller system identification problems. A large number of structure and function measurements need to be made at high resolution.

1.4.1 STRUCTURE

The most promising results in high resolution connectome data are produced through volume microscopy in which electron microscope images are taken at successive ultra-thin layers of brain tissue. In electron micrographs at a resolution of 5-10nm it is possible to identify individual synapses and to reconstruct the 3D geometry of cell bodies of individual neurons with the detailed morphology of axon and dendrite branches. Excellent results have come out of the labs of Winfried Denk (Max Planck), Jeff Lichtman (Harvard) and Ken Hayworth (Janelia Farms). A strong interest in connectome data led to rapid tool development between 2008 and 2011. Two teams used Serial Block Face Scanning Electron Microscopy techniques from the lab of Winfried Denk in combination with two-photon functional recordings and published remarkable results in pieces of retina (Briggman, Helmstaedter, and Denk, 2011), as mentioned above, and visual cortex (Bock et al., 2011). From 3D reconstructions they were able to identify specific neural circuit functions that were corroborated by their functional recordings. This class of tools is well on its way to solving one of the main requirements for whole brain emulation.

1.4.2 FUNCTION

As for the functional data needed from each small sub-system, the most promising tool development is taking its inspiration from the brain's own approach: detection at close range in physical proximity to sources of interaction, namely via microscopic synaptic receptor channels. The brain handles a tremendous quantity of information by utilizing a vast hierarchy of such receptor connections. Similarly, to satisfy the temporal and spatial resolution requirements for *in-vivo* functional characterization, investigative tool development is looking primarily at ways to take the measurements from within.

There is a collaborative effort underway at MIT, Harvard University and Northwestern University to create biological tools that employ DNA amplification as a means to write events onto a molecular "ticker-tape" (Kording, 2011). These have the advantage that they readily operate at cellular and sub-cellular resolutions, and can do so in vast numbers throughout the neural tissue.

Synthetic DNA with a known code is duplicated over and over again through circular amplification. This is done within the cell body of a neuron, but that cell has been modified so that spike events or changes in membrane potential interferes with the amplification process, resulting in a rate of errors that correlates with the activity of the cell. Functional events are thereby recorded on biological media such as DNA. The recordings may then be retrieved from the cells in which they reside.

Another approach is to carry out functional characterization by replacing traditional recording electrodes with micron-scale free-floating wireless probes. Researchers in labs at MIT, Harvard University, UC Berkeley, and other locations are focusing on this approach, tackling issues such as power delivery, communication, probe localization, recording (and stimulating), as well as biocompatibility. In one prototype at UC Berkeley, known as "Neural Dust", free-floating probes contain a Piezoelectric crystal and CMOS circuitry (Seo et al., 2013). Changes of local field potentials in neural tissue are detected and change the resonance frequency of the crystal, which can be queried by ultrasound. Ultrasound, as in information carrier, has the advantage that its energy is not readily absorbed by brain tissue and therefore causes little heating. Another version, conceived in an MIT/Harvard collaboration, investigates a CMOS probe with a possible diameter of 8 micrometers (the size of a red blood cell) that receives power and communicates via infrared light, employing a concept that resembles radio frequency identification (RFID).

Properly developed, technology such as wireless implantable neural probes should be inexpensive, adaptable, accurate and comparatively safe to use, since their application can be less invasive than procedures that break tissue barriers or deliver high doses of electromagnetic radiation.

2. Discussion

The brain's own system components, synapses and neurons are sensitive to information that is conveyed by the temporally specific occurrence of neural action potentials or spikes. That information can be conveyed at rates up to 1 kHz, though usually much less. If this is what the components of a brain can detect then it makes sense that neuroscience tools should be able to record activity data each 1 ms at every neuron (Marblestone et al., 2013). Such tools should then be able to gather the data that enables us to characterize the behavior of a neuronal circuit and to derive functions though system identification.

Using the iterative approach described here, based on rigorous system identification and a decomposition into feasibly characterized sub-systems, a neuroprosthetic reproduction of a mind may be created via whole brain emulation in the coming decades. Some pioneers in the field of artificial general intelligence (AGI) have pointed out areas of overlap between AGI research and neuroscience research that emphasize the value of an interdisciplinary perspective (Goertzel and Pennachin, 2007). The editors of this Special Issue of the Journal of Artificial General Intelligence generally agree with that insight. Clearly, one of the primary causes of interest in AGI has been "to make computers that are similar to the human mind", as Wang (2011) notes unequivocally. Although several AGI researchers are explicitly pursuing forms of (general) intelligence that are designed from first principles without a desire for comparability or compatibility with human intelligence, many approaches and sources of motivation in the search for AGI do involve a strong interest in anthropomorphic interpretations of intelligent behavior.

In past decades, research in AI has been guided by insights about the human mind from experimental and theoretical work in psychology and cognitive science. Insights at that level were the obvious source of information, since very little was known about the underlying mechanistic

architecture and functionality of the brain. For a long time it has been impossible in neuroscience to reconcile the very small with the very large. Investigation at large scale and low resolution was congruent with cognitive science, and led to the identification of centers of the brain responsible for different cognitive tasks through fMRI studies (e.g., Op de Beeck, Haushofer, and Kanwisher, 2008). If we accept that definitions of generality and of intelligence used in AGI can apply to human minds, then a reproduction of the processes of a human mind via whole brain emulation is a type of AGI (Koene, 2012b).

Modeling of thought processes is necessary for whole brain emulation and can be beneficial to efforts in AGI, but the goals and therefore the success criteria are different: AGI is successful if it manages to capture the general principles of a mind to the point where a machine can achieve a desired level of performance for a spectrum of possible tasks. A neuroprosthesis or a whole brain emulation is successful if system identification captures perceived aspects of an individual and personal nature. Due to this difference, there will be points at which the level of investigation in biological brains will be chosen differently to best suit each goal. Another important realization is that both work on artificial intelligence and on WBE are mainly evaluated in terms of performance, and neither necessarily implies a full understanding of human intelligence. That said, whole brain emulation can provide readily accessible working mind functions that may rapidly facilitate insight and understanding.

3. Papers in This Special Issue

There are seven articles in this special issue. The first three offer new algorithmic formulations and approaches to implementation concerning the identification, interpretation and re-implementation of mind functions. The fourth and fifth papers discuss technology forecasting for whole brain emulation, and the last two papers approach the topic of whole brain emulation from the perspectives of legal challenges and risk mitigation.

Sergio Pissanetzky and Felix Lanzalaco propose Causal Mathematical Logic (CML) as a physical explanatory theory that links intelligence with causality and entropy in their paper "Black-box Brain Experiments, Causal Mathematical Logic, and the Thermodynamics of Intelligence". The authors explain that their approach to intelligence is general, so that it may offer a formal link between neuroscience, the emulation of brains and AGI with cross-disciplinary benefits. With CML, Pissanetzky and Lanzalaco consider information processing requirements that must be met to accomplish intelligent operations. Requirements include large causal space, autobiographical memory, and a substrate supporting causal logic. They propose that CML can solve the mind-body problem by quantitatively allowing for explanations in the form of causal associations. Their results include experiments that were carried out using a virtual machine in order to test the application of their intelligence theory. The authors hope to apply further simulations to the discovery of fundamental problems to solve for whole brain emulation. Furthermore, the paper's valuable contributions include an emphasis on the involvement of sentient human perception in the observation of science.

Felix Lanzalaco and Sergio Pissanetzky, in "Causal Mathematical Logic as a guiding framework for the prediction of Intelligence Signals in brain simulation", further expound upon CML, with a general theory of biophysical intelligence, supposing that a large number of observable life and intelligence signals can be described in terms of CML. The paper begins with a useful review of

our understanding of brain systems, then leads to an intuitively pleasing integration of principles of "least action" and "entropic life signals" to explain and predict intelligent perceptual processes identified via event related potentials (ERP) and their constituents.

Leslie G Seymour contributes an insightful proposal for the transformation of LifeLog derived persona specifications into a canonical representation of the neocortex architecture of the human brain, in his paper "Declarative Consciousness for Reconstruction". For a first iteration, the method is described with a good degree of detail and includes a description for the application of incremental compilation technology. The incremental procedure maintains an IT model of the neocortex, which is updated every time novel stimuli are obtained from the ongoing LifeLog. Seymour hopes that the approach can lead to an understanding of the semantic allocation of neocortical capacity.

Daniel Eth, Juan-Carlos Foust and Brandon Whale investigate the plausibility of WBE being developed in the next 50 years (by 2063) in their paper "The Prospects of Whole Brain Emulation within the next Half-Century". The authors carry out a multi-faceted review of requirements for WBE, and they attempt to integrate a number of aspects that were previously not adequately addressed in the literature. Subsequently, they deliver an analysis in terms of possible scenarios and driving forces. Four essential requirements were identified, namely brain scans, translation from scan to model, running the dynamic models, and simulating an environment and body. Among factors that introduced the most uncertainty in the development of WBE, the authors pointed out the need to develop advanced probes that can acquire high-resolution neural data *in-vivo*, as well as the effect of cooperative versus competitive cultures around WBE. The paper makes a good argument for upper and lower bounds on scale-separability among brain mechanisms. Eth *et al* conclude with four scenarios based on the uncertainties, and they suggest a scenario in which WBE is realized and the technology is applied to moderately cooperative ends.

Jeff Alstott, in "Will we hit a wall? Forecasting Bottlenecks to Whole Brain Emulation Development", proposes that a rigorous forecast of the entire technology graph for WBE development is a prerequisite for any forecast of the development of WBE. In the process, we may identify bottlenecks and address them. Alstott points out that most existing forecasts for WBE only consider a fraction of the technology network, mainly studying available computational capacity. Traditionally, such forecasts take an estimate of required capacity and use Moore's law to project when that may be reached. There are more pressing hurdles for WBE, and overcoming those will determine the time-line of progress towards accomplishing whole brain emulation. His paper introduces a framework for describing technology development through technology networks and includes a simple model that illustrates the impact of bottlenecks on forecast accuracy.

Kamil Muzyka considers the legal implications of granting personhood rights to artificial intelligences or emulated human entities in his paper "The outline of personhood law regarding artificial intelligences and emulated human entities". He makes the observation that present-day personhood relies largely on being the offspring of (two) human (genetic material) donors. The paper includes a proposal for a status of "legal adolescence" to be applied in a situation of multiple "selves" (copies) that would allow them to develop into differentiated persons.

Peter Eckersley and Anders Sandberg, in "Is Brain Emulation Dangerous?", assess AGI risks with a focus on human brain emulations, and they discuss the possible fragility of emulation autonomy. They take a good first look at issues of risk surrounding the development of whole

brain emulation, though some aspects introduced by iterative, gradual and piece-wise scientific development exceeded the scope of the paper. Their main conclusion points out that the degree of risk posed by brain emulation probably depends on the order of accomplishments in the research trajectory: Brain emulation may pose fewer risks to society if it is accomplished sooner, because less powerful computers would lead to a more gradual technology impact. Similarly, brain scans produced before a full neuroscientific understanding is reached may result in a larger available initial population of emulations with a better balance of influences when the emulations appear. Eckersley and Sandberg emphasize the connection between WBE risks and the attacker-defender balance of power in computer security challenges. If the allocation of processing power can be regulated then WBE is safer, if processing power can be easily stolen then WBE can more easily lead to destabilizing developments. The authors compare arguments for and against 'open' technology development for WBE and conclude that initial study suggests an open methodology is good policy. Their core conjecture: It is advisable to address neuroscience and microscopy requirements for WBE quickly, in order to reduce the likelihood that emulations appear suddenly and dramatically if there is a surplus of computational capacity.

The work presented in the seven papers covers a broad spectrum of issues surrounding whole brain emulation and the rise of novel forms of intelligence, most of which have not yet been adequately addressed in the research literature. Consequently, these articles will contribute significantly to research in whole brain emulation and to cross-disciplinary efforts.

Acknowledgments

We would like to thank Pei Wang for his indefatigable help in the production of this special issue.

References

Berger, T. W.; Song, D.; Chan, R. H.; Marmarelis, V. Z.; LaCoss, J.; Wills, J.; Hampson, R. E.; Deadwyler, S. A.; and Granacki, J. J. 2012. A hippocampal cognitive prosthesis: multi-input, multi-output nonlinear modeling and VLSI implementation. *IEEE Trans Neural Syst Rehabil Eng* 20(2):198–211.

Bock, D.; Lee, W.-C. A.; Kerlin, A.; Andermann, M.; Hood, G.; Wetzel, A.; Yurgenson, S.; Soucy, E.; Kim, H.; and Reid, R. 2011. Network anatomy and *in vivo* physiology of visual cortical neurons. *Nature* 471:177–182.

Briggman, K.; Helmstaedter, M.; and Denk, W. 2011. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* 471:183–188.

Deca, D. 2012. Available Tools for Whole Brain Emulation. *International Journal of Machine Consciousness* 4:67. doi: 10.1142/S1793843012400045.

Goertzel, B., and Pennachin, C. 2007. *Artificial General Intelligence*. Springer.

Koene, R. 2012a. Experimental Research in Whole Brain Emulation: The Need for Innovative *In-Vivo* Measurement Techniques. *Special Issue of the International Journal of Machine Consciousness* 4(1). doi: 10.1142/S1793843012500047.

Koene, R. 2012b. Toward tractable AGI: Challenges for System Identification in Neural Circuitry. In Bach, J.; Goertzel, B.; and Iklé, M., eds., *Artificial General Intelligence. 5th International Conference, AGI 2012*, 136–147.

Kording, K. 2011. Of Toasters and Molecular Ticker Tapes. *PLoS Computational Biology* 7(12):e1002291. doi:10.1371/journal.cpbi.1002291.

Ljung, L. 2008. Perspectives on system identification. In *In Plenary talk at the proceedings of the 17th IFAC World Congress, Seoul, South Korea*.

Marblestone, A.; Zamft, B.; Maguire, Y.; Shapiro, M.; Cybulski, T.; Glaser, J.; Stranges, B.; Kalhor, R.; Dalrymple, D.; Seo, D.; Alon, E.; Maharbiz, M.; Carmena, J.; Rabaey, J.; Boyden, E.; Church, G.; and Kording, K. 2013. Physical Principles for Scalable Neural Recording. *Frontiers in Computational Neuroscience* 7:137. doi: 10.3389/fncom.2013.00137.

Marmarelis, V. Z.; Shin, D. C.; Song, D.; Hampson, R. E.; Deadwyler, S. A.; and Berger, T. W. 2013. On parsing the neural code in the prefrontal cortex of primates using principal dynamic modes. *J Comput Neurosci* doi: 10.1007/s10827–013–0475–3.

Obama, B. 2013. President Obama Launches the "BRAIN" Initiative. *http://mypro.munawer.in/www.whitehouse.gov/blog/2013/04/02/president-obama-launches-brain-initiative*.

Op de Beeck, H.; Haushofer, J.; and Kanwisher, N. 2008. Interpreting fMRI data: maps, modules and dimensions. *Nature Reviews Neuroscience* 9:123–135.

Seo, D.; Carmena, J.; Rabaey, J.; Alon, E.; and Maharbiz, M. 2013. Neural Dust: An Ultrasonic, Low Power Solution for Chronic Brain-Machine Interfaces. *arXiv:1307.2196*.

Wang, P. 2011. Artificial General Intelligence: A Gentle Introduction. *Retrieved December 31, 2013, from http://sites.google.com/site/narswang/home/agi-introduction*.