

# System Identification: The foundation of modern science and the core of feasible mind uploading

Randal A. Koene

[randal.a.koene@gmail.com](mailto:randal.a.koene@gmail.com)

Carboncopies.org

1087 Mission Street, San Francisco, CA 94103

**Abstract:** We show that every aspect of modern science relies on creating representations of things. And when we do, we pick the signals that interest us and the behavior that interests us. From that, we determine how to interpret the way input is converted into output in a system and our description of that process is our understanding of that system. The same is true for mind uploading, where we have a purpose in mind (to create a substrate-independent version of a specific mind's functions). The feasible approach to this is called whole brain emulation and relies on determining precisely which signals we care about and then breaking the problem down into a collection of smaller system identification problems. To tackle those, we have a roadmap that includes structural scanning (connectomics) as well as new tools for functional recording - some of which I am now working on through a new startup that is collaborating with MIT and Harvard. Using this approach, uploading a mind via whole brain emulation, should become a reality in the next two to four decades.

## Introduction

Tracking down the causes or principal impediments at the heart of many of the problems we face today ultimately boils down to a lack of access to the systems involved, in particular, the systems of the human mind as implemented in the biophysical hardware of the brain. Being able to see what is there, to inspect content, measure or acquire data is primary to the ability to do things like repair, back up, adjust, etc. Having a good idea of what goes on inside our bodies, what the skeletal structure should look like, is extremely useful for reparative surgery. Knowing the design plans of a car helps us replace a broken car. Making copies of valuable documents keeps them from getting lost. We need similar access to the brain. Gaining that access opens up all of those possibilities, and it does not impede or remove any of the quality of our human experience as generated by the processes of the brain.

In recent times, one of the most significant advances in the biological sciences has been the development of tools that allow us to sequence and to synthesize DNA, access to the blueprints of biological form and function. The medical and social implications, the possibilities are as yet hard to fathom, as the field is wide open and largely unexplored. We can compare this with the emergence of universal access to digital and digitizable information with the arrival of the Internet. It is almost impossible now to imagine a world without the Internet in it, without information available from around the world and at our fingertips. When it first appeared, in the form of the Advanced Research Projects Agency Network (ARPANET), it was a useful tool for universities and for the military, but there was no way to anticipate the plethora of applications that followed. Similarly, we have a platform in the form of DNA sequencing and synthesis, but its multitude of uses cannot be predicted – but when we have them, they will become indispensable.

So, we can examine our biological form and we can fix broken bones and even replace organs, or create increasingly powerful prosthetic limbs. We can examine our developmental blueprints in DNA and even apply gene therapy in adulthood (Bernardes de Jesus et al., 2012). The complex structure of the brain has not yet been addressed with tools powerful enough to give us this kind of access. Its structure is initiated by developmental processes guided by DNA and its structure is linked to and influenced by its environment

within a human body, but much of its intricacy is arrived at by a different kind of development: neural plasticity. The resulting operational circuitry is still largely opaque to us. We cannot yet take a snapshot that would enable any sort of backup, would inform repair or reconstruction, the correction of problems or the extension of our abilities and senses.

Understanding how mental functions work requires more than just a rough idea of the locations within a brain that are involved in those functions. It is therefore a frequently unsatisfactory aspect of published studies using magnetic resonance imaging (MRI) that the results rarely give insight into mechanisms specific to a mental operation. Lacking such insight, it is difficult to understand why John responds to an experienced scene in one way, while Mary responds in another. Or, how is it that Mary's brain allows her to concentrate on scientific studies and a passion for exploration, while John seeks primarily distraction and social comforts? What underlies these very individual characteristic differences? Only if we can look at the steps in a mental operation then we can truly understand and make predictions, which is the way in which scientific study leads to understanding. And, only then can we diagnose and treat with confidence, instead of making speculative judgments about mental states and brain health.

Mind uploading to an implementation of mind in a substrate that allows direct insight into the details of ongoing mental functions is the the most complete solution for backup, repair, study and enhancement. By achieving an implementation of intelligence in an engineered processing substrate, a machine mind as it were, this solution is also clearly related to work in artificial intelligence and shares many of its analytical requirements and synthesis goals, even though the objective is unambiguously to make individual human minds independent of a single (biological) substrate (Koene, 2011).

## **Neural Interfaces and Neural Prostheses**

The level of access we need to achieve requires further study and the development of interfaces that can be used long-term for ongoing access to detailed brain activity without causing unwanted disturbances or changes. The same interfaces are needed so that individualized neuroprostheses can be used as clinical replacements for the function of specific brain circuits. The development of brain-machine interfaces has taken several directions, and a significant categorization can be made according to the side of the interface that needs to make most of the adaptations for communication to be possible: the machine or the brain. Seen from that perspective, a large proportion of interfaces today expect the brain to do most of the heavy lifting in terms of learning to adjust and work with the interface, to learn to use it as a channel through which to read out or stimulate (Lebedev & Nicolelis, 2006).

When the brain has to do most of the adapting to the new communication protocol, the interface does not need to understand all the intricacies of neuronal circuitry or how the brain normally interprets signals and produces output through the body. Nor does it need to understand how one brain region speaks with another so that they may together carry out some brain function. In this case, the brain learns to identify and decode the interface and to give the necessary responses, for example, so that a prosthetic limb with limited degrees of motion or feedback can be used. Beyond this practical matter of requiring the interface recipient to carry out the learning and adjusting, it is worthwhile to consider how that adjustment differentially affects identity, personality and the experience of a patient over time, especially when the number of interface channels is drastically increased.

Alternatively, we have to devise interfaces that seamlessly fit into the operating brain. Such interfaces must already be adapted and usable without much learning on the recipient's end, seeming as natural as a person's normal sensations and actions of body and limb. To build such interfaces demands careful analysis of neuronal circuit function. What input does that circuitry receive? What does that circuitry do in terms of its transformations of input into output? The analysis may even need to be carried out in a subject specific manner where the details of circuit function depend on individual characteristic details of the circuitry.

Future developments that may actually accomplish access to the detailed mechanisms and parameters of mental operations will require communication channels with high resolution and high bandwidth. Retraining the brain for each of those channels is not only impractical, but it would alter and distort those same details to which we seek access.

Cognitive prosthetic work, as carried out by the Berger lab at the University of Southern California has already had to address this issue and managed to do so with successful experimental results (Berger et al., 2012). While practical examples of functioning cognitive neural prostheses are still rare, one well-known example is their hippocampal neural prosthetic. The prosthesis depends on a so-called bio-mimetic chip, which implements an algorithmic transfer function that was trained to replicate the operational properties of biological neural circuitry in a region of the hippocampus known as CA3. It manages to copy the way in which input to the region is turned into output from that region. The prosthesis has previously been shown to successfully take over the function of rat CA3 circuitry under experimental conditions and it is presently being tested in primates (Marmarelis et al., 2013).

Brain emulation strives to achieve a mechanistic re-implementation that makes it possible to predict an active state and behavior at a time  $t+dt$  with acceptable error if we know the state at the 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, 1999). By investigating the correlated input and output, you try to determine which functions constitute that characteristic processing. The brain is composed of many similar components, such as neurons of several types and synapses of several types. It is a very large collection of such components and the arrangement is highly complex.

The concrete success of such an approach to brain emulation, a successful neuroprosthesis, is determined with reference to experimental goals or well-define performance requirements that set experiential criteria (Koene, 2012b). One example is the experimental performance of the hippocampal replacement chip, as tested in laboratory settings by Berger's team. Another is the proof-of-concept verification carried out in published work by Brigmann et al (2011), where the connectome of a sample of retinal tissue was studied. They used the structural data obtained by electronmicroscopy in the lab of Winfried Denk to derive functional interpretations for retinal ganglion cells, and the protocol was similar to one that could be used to build model representations with specimen specific parameters for brain emulation. Their publication was an important proof-of-concept, because they were able to verify their functional derivations by comparison with functional data that had been gathered in the same specimen via fluorescent optical microscopy.

## **Representations, Models and System Identification for Mind Uploading**

Every aspect of modern science relies on creating representations of things. And when we do that, we pick the signals that interest us and the behavior that interests us. Understanding is simplifying, abstracting. While abstracting the situation at one level, you seek to connect what is going on at several levels. That creates confidence that you understand how one thing leads to another and what the significant mechanisms are. In turn, that enables prediction, which goes hand in hand with modeling and understanding.

The principle of trying to explain as simply as possible is also known as Occam's Razor. It is a sensible heuristic, because the best way to learn about the actual influences that need to be modeled and understood is to work with a model, learn about its failings and iteratively improve it. The alternative, an exhaustive attempt to include from the beginning all of the possible factors that one imagines might possibly contribute, quickly becomes entirely impractical in any modern field of science.

System identification is a core method of modern science: In science, we can catalog things. We can categorize and create ontologies, and we can make hypotheses about processes. We use representations every

time we make distinctions by measuring or cataloging, and we build models whenever we have mechanistic hypotheses. We use perception as a means of gathering data about our environment and ourselves. And we try to gain the ability to make reliable predictions, i.e. understanding, by testing falsifiable hypotheses. To do this, we have to make sense of data, we need to identify patterns that are discernible within specific constraints. We categorize or measure. Both actions draw boundaries and thereby simplify a problem. The simplification is not meant to ignore reality, but it is meant to acknowledge that there are both deviations (noise) and constrained regularities (main effects) to be expected and therefore to be predicted. We often find that highly simplified, abstract and brief representations, such as  $E=mc^2$  are very powerful means of explaining, understanding and predicting when applied correctly. The functioning of a mind, understanding it in the individual case - or rather being able to make useful predictions based on good representations - can be treated in the same way. The mind is manifested through our physical world and amenable to study by the same careful and diligent methods.

System identification makes mind uploading feasible. To understand this point, consider what is thought to be a possible hurdle to feasible mind uploading? The notion of uploading a mind means that we want to take that which we think embodies or generates ourselves, the feeling of being, the experience of existence - and we want to capture it in a state and move it, or we may want to keep it going and replace the mechanisms that are generating it in-place (Koene, 2012c). We don't want to lose anything we consider important to that experience. Many of the details concerning that experience are still unknown. What we do know is that the mechanisms supporting it are almost certainly contained within the brain and that the precise arrangement is unique for each of us, even though general outlines are similar and repetitive within the brain. The importance of these insights has already been demonstrated through work on the first cognitive neural prosthesis, the hippocampal prosthesis that was developed by Ted Berger's research group. We know that the mechanisms involved appear at the cellular level and that very many operant components cooperate to produce emergent functions.

So, how does a concrete application of system identification overcome hurdles to mind uploading? System identification is applied with regard to the signals that we presuppose convey the operations that are of interest to us. Exploring the realm of candidate signals is a process of iterative hypothesis testing. Beyond this, system identification does not make assumptions about how mental processes are implemented. It is a general approach to systems analysis that is intended to deal with any "black-box" system. We inspect input and output, and derive the transformations (transfer functions) that are necessary to produce the output from the input, transfer functions that can include "memory" (hysteresis) within the system. System identification, as a method, is general enough that we can now concretely discuss systems operating in brain tissue, even develop neural prostheses, while we grow and augment our understanding. In terms of available mathematical formulations and analytical tools, system identification is well developed. We can methodically compartmentalized a large system to describe it as a combination of identified smaller systems. Depending on the experimental measurement tools available, we can adapt system representations and parameters accordingly.

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 (but possibly latent) system behavior. A bigger system with more inputs and outputs will require that a larger number of cases be observed. So, how big can an unknown system be for which we can reliably identify functions? If we chose the entire brain as a single unknown system then we would probably have to observe its input and output throughout its entire life-span. What we could deduce from the resulting data would still be flawed and would likely fail to capture latent functions. Furthermore, successfully tuning any emulation attempted in that way would be computationally intractable when the brain contains billions of neurons and trillions of synapses. Instead, we can break the problem down into smaller pieces, sub-systems that communicate with one-another. We would like to end up with a collection of individually manageable system identification problems that fit our ability to carry out measurements, collect data, model functions, estimate parameters and ultimately devise prosthetic replacements.

We need to do three things: 1.) We need to choose those smaller sub-systems. 2.) We need to know how they are connected and communicating. And, 3.) we need to make measurements at each sub-system to identify its system functions. Considering the problems and possible solutions for those three needs of system identification helps us build a roadmap toward emulating functions of brain tissue.

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 in the roadmap toward this so-called Whole Brain Emulation according to 4 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;
4. Emulation - the mathematical representations and computational platforms needed.

Whole brain emulation is an absolutely essential goal for neuroscience: To truly understand a thing you must be able to build it. To build it we need models, a combination of building blocks with processes. When we explain something (e.g. mental functions and behaviors) we make it 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 outcome prediction. Within the defined system outlines of our model, taking into account defined sets of signals, we describe interactions (which we may formally express in information theoretic terms).

We do not need to pretend that we know immediately what all the proper contributors are or that we fully understand how to accurately make system predictions. We can formally discuss mental processes without being restrictive about the way in which they may be achieved. Coming up with a satisfactory model or theory is an iterative process based both on conceptualization and on data evaluation, and we may make some initial assumptions based on the presence or lack of some 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/replicate the processes of the mind. Once we have a model with processes acting on signals and we acknowledge that boundaries are drawn around systems in some way then, when we focus on a specific set of (sub-)systems, we can identify the boundary between them. For example, 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 within the systems we focus on, for example, experiences attended to versus undesired/randomized/other experiences, spatially constrained sub-systems (e.g. 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 system (including hysteresis/memory in the system). True to our purpose, we determine how to interpret the conversion of input into output and our description of that system process becomes our understanding of the system. The same is true for mind uploading, where our purpose is to create a substrate-independent mind (Koene, 2012b).

The aim is to be explicit about our interests and success criteria (concepts of satisfactory accomplishment). That applies to the replication of specific mental operations, e.g., the ability to acquire new episodic memory that was replicated in the hippocampal neuroprosthesis devised by Berger. It also applies to the replication of the full range of mental operations, the process of “mind uploading” into prostheses for all characteristic mental processes. Without being explicit, it would be impossible to carry out the iterative exploration of

model spaces, even though that exploration itself further refines our goals. Processes described and parameter states established within a representation become a substrate-independent version of a mind, in accordance with the goal criteria. We can use any means necessary and available to carry out predictions, i.e. to express signal exchanges, to generate experiences. When we generate experiences by predicting plausible and probable signal exchanges at the level of neural biophysical processes in complete volumes of brain tissue then we are applying an emulation approach with regard to the biological brain: whole brain emulation.

## **Iterative Improvements in Four Main Areas to Achieve Whole Brain Emulation**

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. To tackle those, we have a roadmap that includes structural scanning (connectomics) as well as new tools for functional recording.

The method that is producing the most promising results in terms of connectome data today is high resolution volume microscopy carried out by taking electron microscope images at successive ultra-thin layers of brain tissue. Electron micrographs at a resolution of 5-10nm enable us to identify individual synapses and to reconstruct the 3D geometry of fine detail in the axon and dendrite branches. Excellent results have come out of the labs of Winfried Denk (Max Planck), Jeff Lichtman (Harvard) and Ken Hayworth (Jenalia 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 that may be proof-of-principle for system identification in pieces of retina (mentioned above) and visual cortex. From 3D reconstructions they were able to identify specific neural circuit functions that were corroborated by their functional recordings, such as for the direction selectivity of retinal ganglion cells (Brigmann et al., 2011). 3D reconstructions clearly show the cell bodies of individual neurons, and the detailed morphology of axons and dendrites, situated within neuronal circuitry with visible synaptic connections. It is important to realize that this is no longer just an idea for future developments, but a class of existing tools that can directly solve one of the main requirements for whole brain emulation.

As for the functional data that is required for system identification in each small sub-system, the most promising developments take their inspiration from the brain's own method to detect activity: at very close range in physical proximity to sources of interaction, namely through microscopic synaptic receptor channels. The brain handles an enormous quantity of information by using a vast hierarchy of such connections. So, to satisfy the temporal and spatial resolution requirements for in-vivo functional characterization we look primarily to the development of new tools that can take these measurements from within.

There is a collaborative effort underway at MIT, Harvard 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 also received a voltage-gated channel that 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. The project goal is to be able to record signals from all neurons in a brain, and potentially to measure at resolutions beyond that.

Another approach is to carry out functional characterization by establishing the equivalent of an electronic synaptic network based on micron-scale wireless neural interfaces. Researchers in labs at MIT, Harvard, UC Berkeley, and other locations are focusing on this approach, tackling the various issues such as power, communication, localization, recording, stimulating, biocompatibility. In one prototype technology under active development at UC Berkeley, so-called “Neural Dust” is composed of free-floating probes that contain

a Piezoelectric crystal within a capacitive loop (Seo et al, 2013). Changes of local field potentials in neural tissue will change the resonance frequency of the crystal, which can be detected by probing the crystals with ultrasound. In another prototype technology envisioned by an MIT/Harvard collaboration, a probe with a diameter of 8 micrometers, the size of a red blood cell, can include operational circuitry, infrared power delivery and communications, and an antenna for passive communications similar to radio frequency identification (RFID). An infrared version of this passive communication, a micro-OPID (at optical wavelengths), is being developed by Dr. Yael Maguire. Using 32 nanometer IC technology, the probe that can fit into capillaries of the brain vasculature that supply every neuron can have 2300 or more transistors. 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. Probes at micron scales can be self-contained, requiring minimal assembly. Based on integrated circuit (IC) technology, the trajectory of their development and production can take advantage of Moore's Law.

When we have the tools and make all the necessary measurements then we end up with a vast amount of data such as brain images and recordings of voltage changes near neurons. How do you turn that data back into a functioning mind? We need ways to generate models and to populate those models with parameters based on the pool of measurements. Measurements of different types need to be combined, such as structure data and activity data. We also need computing platforms that are well-suited to an implementation of the emulation. Some work in this direction has been done in studies of efficiently distributing volumetric neural tissue computations over many nodes of a supercomputer (Kozloski & Wagner, 2011).

Brain emulation on general purpose computers is convenient, because model functions can be modified easily. Ultimately, a mature emulation demands a suitable computing substrate. We know that low power solutions that fit within 10 cubic centimeters are possible in principle, because those are the specifications achieved by the biological brain. There, the functions of mind are carried out by a vast network of mostly inactive, low-power processors: neurons. A new substrate with similar features may achieve efficient emulation, and the development of so-called neuromorphic computing platforms points in that general direction. Examples are: Chip designs resulting from the DARPA SyNAPSE project, outcomes of the EU CAVIAR and FACETS projects, and extensible neural microchip architectures such as developed by Guy Paillet.

Iterative improvement of our understanding of the system identification problem seeks to converge effectively on methods that satisfy our success criteria. Can we make do with a simple model of neurons, or might we even need to model molecular kinetics explicitly? Must we take into account specifics of each glial cell in addition to the neurons? The more we know about the relevant signals and the architecture of the brain, the smaller we can make the sub-systems for which system identification needs to be carried out. Determining the right scope and resolution for emulation is an outcome of iterative hypothesis testing using our models and data acquisition tools.

Theoretical debates occasionally focus on the question of determinism and computation in the Von Neumann sense. Is it impossible to initially implement a successful whole brain emulation in terms of a software program within a digital computer, because some aspect of the operation of biophysical primitives in the brain may be non-deterministic? We can say that neurons and other parts of the physiology are not deterministic and therefore not like computer programs. We could say the same thing for anything else that is built in the real world, such as transistors. Notice that transistors also do not operate in a totally deterministic and predictable manner. They too are subject to the whims of variability in material, effects from surrounding environmental noise. At the highest resolution, the transistors within a modern CPU are not all identical and they will not all exhibit exactly the same response function. Additionally, they will be affected by many other unpredictable quantities, such as the surrounding temperature, magnetic fields and other sources of radiation. From this, we might deduce that given the hardware of a computer it cannot possibly be a deterministic machine that could be represented as running programs. At high resolution, all the building blocks of nature and our world are subject to variability that appears non-deterministic at least at the scale where mesoscopic

and macroscopic processes need to take place reliably.

Of course, we know that engineers go to great lengths to make the components of a CPU work together in a reliable manner. This includes using a clock for synchronous timing, using thresholds and binary encoding instead of analog representations on the transistors, and including error detection and error correction technology in processes and memory buses. But how does a biological system operate to carry out things like goal-directed behavior or reliable responses if it were truly non-deterministic? The mind needs to operate in a predictable, deterministic manner, even if its components do not. The brain goes to great lengths to make its mesoscopic and macroscopic operations more predictable, more deterministic. There are numerous ways to get around physiological unreliability, to make reliable communication and collaboration between brain regions and circuits possible. Of course, the biological implementation of those solutions is directly related to the identification of the appropriate level of representation and parameter measurement for brain emulation.

Natural selection resulted in solutions that would reliably work, even when depending on billions or trillions of components within many different operational regions that need to collaborate and upon which the survival of the organism can depend. Those solutions can be quite different than the ones in the CPU example, but they are quite rigorous and effective. Examples of measures taken within the brain include:

- Region- or brain-wide modulation of membrane potentials in accordance with large-scale rhythms in the brain (such as the ones detectable by EEG). This effectively gives the brain a synchronizing clock, or rather several clocks running at related oscillation frequencies.
- The use of ensembles for encoding, not relying on the firing of a single synapse or even of a single neuron.
- The use of powerful signals, such as bursts of hundreds or thousands of neural spikes.
- The use of rates of spiking in some places instead of individual spike times (although individual spike timing is meaningful in some brain functions).
- Large-scale modulators that can control the operating modes and conditions of regions or the entire brain.

And the measures include requiring repeated processes to take place before changes are made in the system:

- A single pre- and post-synaptic spike is not usually sufficient to cause a lasting and significant change in a synapse, but several repetitions of the same are.
- A tiered system of memory that will elevate an experience or learning to permanence only if it passes a number of thresholds. Most of what is captured temporarily by iconic or short term memory is lost forever.

Ultimately, you end up with a system that is far more predictable than its fundamental building blocks. The effects that noise and non-determinism do have in the system certainly may be characteristic of the system, but even then it is possible to engineer noise sources with very similar characteristics. There is no indication that satisfactory replication in accordance with criteria for whole brain emulation would be greatly affected by a need to replicate non-deterministic noise.

## **A Platform for Iteration Between Model and Measurement**

ARPANET first appeared in 1969 and developed throughout the 70s and 80s. It was useful for university projects and for the military, but it was not immediately obvious what sort of civilian applications would thrive on such a computer network. The Internet was its outcome, and the every corner of the world was changed greatly as a result. Right now, a transition is underway from simple neural interfaces to interfaces with greater bandwidth and with the ability to interpret or directly use the signals the brain is producing, rather than training the brain to use the interface. In fact, research goals include the development of probes that can record at 1ms sample rates from every neuron in a piece of tissue or an entire brain. There is a resemblance with the situation in 1969: We cannot yet anticipate all of the applications, devices and possibilities that may emerge, but it is clear that there is great opportunity to build the platforms for a new type

of communication and integration with the mind. My own startup, NeuraLink is engaged in this effort, to take the most promising new neural interfaces technologies that are coming out of the lab in places like Harvard, MIT and UC Berkeley, and to make those reliable, easy to use and easy to produce. The task is to build the backbone for the next technological revolution.

Many signal modalities are left to explore and effective hybrid systems can be built. Some of the most exciting recent developments came out of a summit at Harvard on June 12, 2013, which gathered leading neural engineers with the stated goal to identify technologies that could sample activity data at every neuron in a brain at 1ms resolution in-vivo (Marblestone et al, 2013). That goal is a milestone in the roadmap to whole brain emulation, and the concept of whole brain emulation was acknowledged and recognized as a valid and desirable research target during the special Brain Researchers' Meeting that preceded the Global Future 2045 Congress in New York City in June of 2013.

The brain's own system components, synapses and neurons are sensitive to information that is conveyed by the appearance at specific times of spike signals. That information can be conveyed at rates up to 1kHz (though often at lower rates). If this is what the components of a brain can listen to, then it makes sense that neuroscience tools aimed at characterizing the behavior of a neuronal circuit, at devising its functions through system identification, or aimed at co-existing with neuronal circuitry for the purposes of neural interfacing should be able to record activity data at time intervals of 1ms. The problem of system identification and of interacting with the distributed activity of brain systems becomes more tractable when such recording is carried out at more and smaller pieces of the circuitry, for example, at every neuron.

## **Conclusions – Uploaded and Machine Minds**

Using the iterative approach described here, based on rigorous system identification and a decomposition into feasibly characterized sub-systems, uploading a mind via whole brain emulation can become a reality in the next two to four decades. I share with several of the pioneers of artificial general intelligence (AGI) the conviction that there are areas of overlap between AGI research and neuroscience research in which an interdisciplinary perspective is of particular value (Goertzel & Pennachin, 2007). Of course, one of the primary reasons for interest in AGI (and artificial intelligence or machine minds in general) has been “to make computers that are similar to the human mind”, as Pei Wang notes unequivocally (in *Artificial General Intelligence: A Gentle Introduction*, <http://sites.google.com/site/narswang/home/agi-introduction>). Some AGI researchers are explicitly pursuing forms of (general) intelligence designed from first principles and without a desire for comparability or compatibility with human intelligence. By and large though, many of the underlying objectives that drive the search for AGI also involve an interest in anthropomorphic interpretations of intelligent behavior. The operations of the human mind are of interest to strong AI and many in the field of AGI.

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 Beek et al, 2008). Having a rough map of the localization of the known set of gross cognitive functions within a standardized brain does not actually add significantly to an understanding of exactly how the brain does what it does. If we accept that definitions of generality and of intelligence used in AGI can apply to human minds, then a substrate-independent implementation of a human mind is an artificial version of the necessary functions. That makes the substrate-independent mind a type of AGI (Koene, 2012). Emulation is a concrete approach to the transition from a biological brain to implementation in another substrate and it is reasonable to assume that the result will be a functioning mind with general intelligence.

Modeling of thought processes is necessary both in artificial intelligence work and when creating

neuroprostheses for a whole brain emulation. But, the goals and therefore the success criteria are different: AI 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 are successful if they capture perceived aspects of an individual and personal nature. Due to this difference, there will be points at which different levels need to be chosen for system identification in existing brains.

In the preceding, we explained that modern science relies on creating representations of things, where we pick the signals that interest us and the behavior that interests us and then determine how to interpret the manner in which system input is converted into output. Our description of that process is our understanding of that system. For practical purposes, a similar approach applies to mind uploading, where we seek to create a substrate-independent version of a specific mind's functions. To achieve that via 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. We showed that a roadmap to whole brain emulation includes structural scanning (connectomics) as well as new tools for functional recording.

Within 5 years, from 2008 through 2012, we saw connectomics grow from a call for novel research and development to a solid field with proof-of-principle results and sophisticated new tools. In 2013, we have the Human Brain Project. Now large scale high resolution activity mapping has emerged as the new challenge in neurotechnology, and this was made abundantly clear in aims of the BRAIN Initiative. In another 5 years, by 2018, we may have methods for both large scale high resolution structural and functional data acquisition. A feasible project proposal may then involve the analysis and emulation of the neural tissue functions of the fruit fly *Drosophila* – the primary animal studied at Janelia Farms Labs. Emulation of a *Drosophila* brain constitutes a breakthrough of incredible proportions, because it is the brain of an animal that shares with us so many basic features: mobility, learning, planning, sensation. Such an accomplishment will effectively make mind uploading a reality.

## References

Berger TW, Song D, Chan RH, Marmarelis VZ, LaCoss J, Wills J, Hampson RE, Deadwyler SA, Granacki JJ. (2012). A hippocampal cognitive prosthesis: multi-input, multi-output nonlinear modeling and VLSI implementation. *IEEE Trans Neural Syst Rehabil Eng.* 2012 Mar;20(2):198-211. doi: 10.1109/TNSRE.2012.2189133.

Bernardes de Jesus, B., Vera, E., Schneeberger, K., Tejera, A. M., Ayuso, E., Bosch, F. and Blasco, M. A. (2012), Telomerase gene therapy in adult and old mice delays aging and increases longevity without increasing cancer. *EMBO Mol Med*, 4: 691–704. doi: 10.1002/emmm.201200245

Briggman, K., Helmstaedter, M. and Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina, *Nature*, Vol. 471, pp. 183-188.

Goertzel, B., Pennachin, C. (2007). *Artificial General Intelligence*. Springer, New York.

Koene, R.A. (2011). AGI and Neuroscience: Open Sourcing the Brain. *Proceedings of the Fourth Conference on Artificial General Intelligence (AGI2011)*. August, 2011. Mountain View, CA.

Koene, R.A. (2012a). Toward Tractable AGI: Challenges for System Identification in Neural Circuitry. *Artificial General Intelligence, Lecture Notes in Computer Science Vol. 7716*, pp 136-147.

Koene, R.A. (2012b). Experimental Research in Whole Brain Emulation: The Need for Innovative In-Vivo Measurement Techniques. *Special Issue of the International Journal on Machine Consciousness*. Vol.4(1),

doi: 10.1142/S1793843012500047.

Koene, R.A. (2012c). Mind transfer: human brains in different materials. *New Scientist*, Issue 2888, November 2, 2012.

Kording, K.P. (2011). Of toasters and molecular ticker tapes. *PLoS Computational Biology*, Vol. 7(12): e1002291. doi:10.1371/journal.pcbi.1002291.

Kozloski, J. and Wagner, J. (2011). An Ultrascalable Solution to Large-scale Neural Tissue Simulation, *Frontiers in Neuroinformatics*, Vol. 5, No.15, doi: 10.3389/fninf.2011.00015.

Lebedev M.A., Nicolelis M.A. (2006) Brain-machine interfaces: past, present and future. *Trends Neurosci.*, 29: 536-546.

Ljung, L. (1999). *System Identification — Theory For the User*, 2nd ed. PTR Prentice Hall, Upper Saddle River, N.J.

Marblestone, A.H., Bradley M. Zamft, Yael G. Maguire, Mikhail G. Shapiro, Thaddeus R. Cybulski, Joshua I. Glaser, Ben Stranges, Reza Kalhor, David A. Dalrymple, Dongjin Seo, Elad Alon, Michel M. Maharbiz, Jose Carmena, Jan Rabaey, Edward S. Boyden, George M. Church, Konrad P. Kording. (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., Berger, T.W. (2013). On parsing the neural code in the prefrontal cortex of primates using principal dynamic modes. *Journal of Computational Neuroscience*. August 2013; doi: 10.1007/s10827-013-0475-3

Op de Beek, H.P., Haushofer, J., Kanwisher, N.G. (2008). Interpreting fMRI data: maps, modules and dimensions. *Nature Reviews Neuroscience*, vol. 9, pp. 123-135.

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