NEOBASE
Databasing the Neocortical Microcircuit

Asif Jan Muhammad a, 1 , Henry Markram a

a Brain Mind Institute, Ecole Polytechnique Fédérale De Lausanne (EPFL), Switzerland

Abstract: Mammals adapt to a rapidly changing world because of the sophisticated perceptual and cognitive function enabled by the neocortex. The neocortex, which has expanded to constitute nearly 80% of the human brain seems to have arisen from repeated duplication of a stereotypical template of neurons and synaptic circuits with subtle specializations in different brain regions and species. Determining the design and function of this microcircuitry is therefore of paramount importance to understanding normal and abnormal higher brain function. Recent advances in recording synaptically-coupled neurons has allowed rapid dissection of the neocortical microcircuitry thus yielding a massive amount of quantitative anatomical, electrical and gene expression data on the neurons and the synaptic circuits that connect the neurons. Due to the availability of the above mentioned data, it has now become imperative to database the neurons of the microcircuit and their synaptic connections. The NEOBASE project, aims to archive the neocortical microcircuit data in a manner that facilitates development of advanced data mining applications, statistical and bioinformatics analyses tools, custom microcircuit builders, and visualization and simulation applications. The database architecture is based on ROOT, a software environment that allows the construction of an object oriented database with numerous relational capabilities. The proposed architecture allows construction of a database that closely mimics the architecture of the real microcircuit, which facilitates the interface with virtually any application, allows for data format evolution, and aims for full interoperability with other databases. NEOBASE will provide an important resource and research tool for studying the microcircuit basis of normal and abnormal neocortical function. The database will be available to local as well as remote users using Grid based tools and technologies.

Keywords: Databases, Data Management, Neocortical Microcircuit, Distributed Applications

Introduction

The neocortical microcircuit is unique in that neocortical neurons are arranged in layers (layers I-VI) that connect to different cortical and sub-cortical regions (Jones,

1 Corresponding Author: Brain Mind Institute, Ecole Polytechnique Fédérale De Lausanne (EPFL), Switzerland; Email: asif.jan@epfl.ch
In the horizontal dimension, the neocortex is functionally parcellated into collaborative groups of neurons, commonly thought of as functional columns (Mountcastle, 1957; Hubel & Wiesel, 1962). In rodents, a neocortical column of about 0.3mm in diameter contains roughly 7500 neurons (100 neurons in Layer I; 2150 in Layer II/III; 1500 in Layer IV; 1250 in Layer V and 2500 in Layer VI). Most neocortical neurons (70-80%) are excitatory pyramidal neurons (Peters & Jones, 1984; White, 1989; DeFelipe & Farinas, 1992), which have relatively stereotyped anatomical, physiological and molecular properties (DeFelipe & Farinas, 1992; Toledo-Rodriguez et al., 2003). The remaining 20-30% of neocortical neurons are interneurons, mostly inhibitory interneurons, which have extremely diverse morphological, physiological and molecular characteristics (Houser et al., 1984; Peters & Jones, 1984; White, 1989; DeFelipe, 1993; Cauil et al., 1997; Kawaguchi & Kubota, 1997; DeFelipe, 2002; Toledo-Rodriguez et al., 2003).

Our laboratory has recorded from over 1000 neocortical neurons, we have reconstructed over 500 of these neurons, and over 200 synaptic connections between specific neurons. We have also studied the expression of 50 genes in over 200 neurons. We are now accumulating data on the microcircuit at an even greater pace and estimate that we will have recorded from over 3000 neurons by 2008, which will provide a realistic sample of the different types of neurons and all major pathways in the neocortical microcircircuit. We have also established industrial scale 3D computer reconstruction facilities which will yield 3D models and detailed morphometric data on virtually all the neurons. We have also developed a gene expression protocol with femtogram-sensitivity for mRNA which enables single cell DNA microarray analysis and we have obtained routine access to a high throughput Affymetrix DNA microarray facility. We therefore aim to obtain the expression profile for all major cell types in the neocortical microcircuit within the next 3 years. In addition, progressively more labs are obtaining quantitative data on the neocortical microcircuitry that will add considerably to the volume and diversity of the data.

It has now become imperative to database the microcircuit components and their connections. A unique style of database is however required to allow databasing of microcircuit data not only for archiving, but also for Neuroinformatics research, for building of custom versions of the microcircuit, for visualization of individual neurons, small groups of cells or the entire microcircuit, and for simulating the microcircuit at various levels of detail. A microcircuit database needs to be specific to the type of microcircuit studied because of the specific arrangement of and connections between neurons. We had earlier constructed a prototype database (Markram et al, 2003) to expose and work through the important issues of databasing microcircuit information. The knowledge and experience from the prototype was utilized for designing and building a futuristic platform, called NEOBASE, for storing the microcircuit data. The NEOBASE will organize the microcircuit data for optimal storage, knowledge sharing, analysis, and visualization and future simulations.
1. Data Model for Neurons and Synapses

Neurons are characterized in terms of their morphological, physiological and gene expression profiles and synaptic connections are characterized in terms of their morphological and physiological profiles. Neuron morphology profiles are obtained from a detailed morphometric breakdown of 3D-reconstructed neurons (m-Profiles), neuron physiology profiles are obtained from detailed measurements of the electrophysiological responses to a series of stimulus protocols (e-Profiles), and neuron gene expression profiles are obtained from single cell RT-PCR and DNA microarray data (g-Profiles). Synaptic connections are characterized by the identity of the pre- and postsynaptic neurons (sn-profile), the axonal and dendritic location of light microscopically identified putative synapses (sm-Profile), and the physiology of synaptic connections obtained from recording postsynaptic responses to a series of stimulation protocols applied to the presynaptic neuron (se-Profile). (Markram et al, 2003) provides a detailed treatment of the various Neuron and Synapse profiles, these profiles are briefly described in the following paragraphs.

1.1. Neuron Profiles

The Neuron data is classified into Morphology, Electrophysiology and Gene Expression profiles. The following paragraphs describe various Neuron profiles.

1.1.1. Morphology Data (m-Profile)

To obtain the m-Profile for a neuron, the neuron is fully 3D-computer reconstructed and converted into Neurolucida format (Glaser and Glaser, 1990). This 3D-model is then uploaded into the database and a MATLAB-based tool automatically performs an extensive morphometric analysis on the model neuron and enters the m-Profile into the database. The m-Profile is a vector of more than 200 values that represent various aspects of the geometry of the neuron. Examples of m-Profile data include TreeLengthMean (mean of lengths of segments with same order in each tree), IndivTreeLengthMean (mean of segment lengths in a tree), and XY_Angle (angle between projection of a segment on XY plane and X axis).

1.1.2. Electrophysiology Data (e-Profile)

The electrophysiological profile of neurons is obtained by applying a series of different wave forms of current injection into the somata of a neuron, during intracellular or whole-cell patch-clamp recording. The responses to these pulses are measured to obtain a spectrum of electrophysiological parameters (EPs). Examples of EPs include action potential active properties (such as action potential waveform, after action potential, and discharge parameters) as well as passive properties (such as input resistance, membrane rectification, and membrane time constants).

1.1.3. Gene Expression Data (g-Profile)

The genetic profiles is obtained from single cell multiplex RT-PCR studies (non-quantitative and quantitative) and single cell DNA microarray analyses. The PCR studies that have been carried out so far enabled non-quantitative detection of the express vs non-expression of 50 genes (Toledo-Rodriguez et al., 2004). The g-Profile is divided into functionally characterized groups of genes such as calcium binding
proteins, neuropeptides, neurotransmitter enzymes, structural proteins and more (Cauli et al., 1997; Toledo-Rodriguez et al., 2004).

1.2. Synaptic Profiles

Multineuron recording allows simultaneous characterization the anatomy and physiology of synaptic connections between identified pre and postsynaptic neurons. The data collected about synapses is organized in the following profiles:

1.2.1. Morphology Data (sm-Profile)

The anatomy of a synaptic connection is described by the sm-Profile. This profile contains information about the numbers of putative synapses, their location on the axonal and dendritic arbors of the pre- and postsynaptic neuron, respectively (also referred to as Synaptic Innervation Patterns), and the axonal and dendritic geometric and electrotonic distances of each of the putative synapses (see (Markram et al., 1997; Wang et al., 2002)). The examples of sm-Profile parameters include Axonal Branch Order (number of branch points between the bouton forming the synapse and the soma of the source neuron), Dendritic Branch Order (the location of the synapse along the dendritic arbor according to the branching frequency of the dendritic tree, and Geometrical Distance (the distance along the dendritic from the synaptic location to the postsynaptic soma).

1.2.2. Electrophysiology Data (se-Profile)

The electrophysiological properties of synapses are characterized in terms of the biophysical, quantal & dynamic properties (Markram et al., 1997; Gupta et al., 2000). The biophysical properties focus on the amplitudes, latencies, rise and decay times of PSPs and/or PSCs; synaptic conductances; synaptic charge transfer, etc. The quantal parameters include estimates of quantal size, probability of release and number of functional release sites. The dynamic properties include the time-constants governing the rates of recovery from synaptic depression (D) and facilitation (F) as well as the absolute and effective utilization of synaptic efficacy parameters.

1.2.3. Pharmacological Data (sp-Profile)

The pharmacological properties are described in terms of their responses to various blockers, agonists and antagonists. Examples for commonly used chemicals are bicuculine (GABA-a antagonist), APV (NMDA receptor antagonist), CNQX (AMPA receptor antagonist), CGP 35348 (GABA-b antagonist), NMDA (NMDA receptor agonist), diazepam (GABA-a facilitator). The sp-Profile contains information describing the sensitivity of the synaptic connection to the different chemicals, and at which concentration.

1.3. Additional Profiles

In addition to storing data about various Neuron and Synapse profiles as discussed above, the NEOBASE will also record Neuron Models and Canonical Data arranged in the following profiles:
1.3.1. Model Data (mod-Profile)

The database will allow for the depositing NEURON (Hines, M 1994) models of each neuron. The NEURON model will include active properties by inclusion of ion channel constellations and parameters. Electrical properties of the neurons could be used in target functions to derive the optimal parameter settings to reproduce the electrical behavior of the neuron. Possible ion channel constellations could also be constrained by the ion channel genes that are found to be expressed by the different neurons. This section will therefore contain a complete model neuron for download and the mod-Profile will contain the values for the parameters of the model.

1.3.2. Canonical Data (x-Profile)

As the database becomes heavily populated, the mean statistical properties of each neuron will be used to build canonical neurons for each type of neuron and each type of synaptic connection. The canonical data section will contain all information as for individual neurons, including images, traces and all the neuronal and synaptic profiles. The section could also contain various degrees of simplification of the neuronal axonal and dendritic arborizations. mod-Profiles of canonical neurons will also be generated.

The following paragraphs describe the illustrate the schematic layout of the NEOBASE, describe various system components, and present the current status of the project.

2. Implementation Details

The following paragraphs describe the implementation details for the NEOBASE project.

2.1. Database Platform

Databasing the microcircuit data requires a highly flexible platform that permits construction of an architecture which allows the data to be mapped, analyzed, and visualized as a circuit. One such highly flexible platform is provided by the ROOT system (Rene Brun and Fons Rademakers, 1996). ROOT is a highly evolved open source software environment built over many years at the European Organization for Nuclear Research (CERN), Switzerland. ROOT enables the construction of an object-oriented database with limitless relational capabilities, thus making it an ideal platform for databasing the microcircuit. ROOT consists of large number of elementary “classes” (currently over 300) that have been contributed by different researchers and allows for the construction of new custom classes as required by different applications and programs. A class is the blueprint for a specific behavior and for a list of attributes, but it has no existence on its own – analogous to an empty form that needs to be filled in. Once information is entered into the class, the class will behave accordingly and will contain the attributes entered. Entering of data in a class is referred as object instantiation i.e. an object corresponding to the data values is created and it captures all

---

2 In the software terminology, classes that describe the behavior of some commonly applicable operations are sometimes declared as “static” classes, in that case the class provides the support for all operation and there are no objects that belong to this class.
the attributes as described by the class and provides implementation for the behavior, in order to complete the desired task. However, objects in the object-oriented programming language can be arbitrarily complex, as they can contain not only a set of attributes, but also a set of operations to perform under various conditions. Additionally, they also form relationships with other objects according to specified rules. A collection of these objects and their interrelationships can be used to construct an object-oriented database which closely mimics the biological architecture. Capability of the objects to relate to each other is an important feature that will allow dynamic 3D reconstruction of thousands of neurons each contacting a specific part of a particular neuron. Furthermore, higher level classes will be designed to contain multiple objects which can be used to group, for example, neurons in different layers of the neocortex.

The following figure provides the schematic layout of the NEOBASE system. As seen from the figure; the NEOBASE consists of three layers i.e. database layer, Web Portal and Web Services layer, and data access layer. The subsequent paragraphs describe the architectural as well as implementation details of these layers.

![Schematic layout of the NEOBASE system](image)

**Figure 1: Schematic layout of the NEOBASE system**

### 2.2. Database Layer

In NEOBASE, we will model all elementary entities of the microcircuit, such as a specific gene, ion channel, or receptor as a class, called, for example GeneX, IonChannelX, ReceptorX, NeuroTransmitterX. These classes will allow the storage of multidimensional data. For example, the IonChannelX class will allow the storage of the gene object that makes the ion channel (constructed with GeneX), the amino acid sequence of the ion channel, information about the density and distribution of the ion channel, information about the ion channel kinetics and even a computer model of the
ion channel. All genes, ion channels and receptors will be grouped into higher level
classes called Genes, IonChannels, Receptors and Neurotransmittters respectively.
Genes will be used to database all the genes found in a particular neuron and this
collection of genes will be stored as an object. A neuron will be databased using a high
level class called NeuronX. NeuronX will contain the objects belonging to Genes,
IonChannels, Receptors, Neurotransmitters, MorphologyX and ElectricalX,
ConnectionX classes. NeuronX can even contain an entire NEURON model of the
neuron generated with NeuronModX. When the different classes are filled in for a
particular neuron, the neuron will be intuitively composed of a specific profile of
objects at various levels and various dimensions. Section 3 of the current documents
briefly describes the current version of the database schema.

2.2.1. Database Schema Evolution

The ROOT system supports the notion of the schema evolution i.e. where the
description of the object classes is changed during the course of time. This allows the
system to query and use the objects described using a combinations of old and new
class definitions. The concept extends beyond reading simple objects and supports
more complex situations i.e. objects with multiple levels of inheritance, as well. The
schema evolution is achieved in the following manner. With every object instance, the
description of the class is stored in the dictionary objects. The dictionary contains,
amongst other parameters, the version number for all the classes in the hierarchy.
When reading a version of the object, its class definition is also read. This definition is
then compared to the in-memory definition of the same class. In case there is a
mismatch between the in-memory version of the class and the persistent version of the
class, then the persistent definition is mapped to the in-memory definition. And the
ROOT system takes care of changing the order of the data members, deleting the old
members, adding new members etc. Thus, it will always be possible to read often
inconsistent versions of same objects and differentiate their usage based on their
version information and associated class definition.

2.2.2. Data Querying, Update and Retrieval

Support for simple as well as complex search, update and data retrieval queries is
one of the important characteristics of any database system. Additionally, the database
system shall also assist the users in discovering the database metadata properties in
order to construct meaningful single and multi attribute queries. The NEOBASE
system will provide support for discovering the metadata attribute information for
various objects in the database, allow for execution of multi attribute search queries,
and provide efficient mechanisms for data update and retrieval. Special consideration
has been given to the efficient execution of search operation in terms of reducing the
storage requirements and improving the query response time. The ROOT framework
provides, through ROOT trees, an optimized search strategy for querying large number
of complex objects. For example, if we have large number of neuron objects, say 1
million of them, and we want to find neurons that have anatomical type as Pyramidal
Cell (PC) belong to layer 2 and have total length of dendrite and axonal trees more than
300 micro meters. Loading each neuron object in the memory and checking its
attributes for above mentioned values will have huge memory and processing
requirements and will not be very efficient. Executing the above mentioned query in
the ROOT will result in loading the individual attributes of the objects (as against to
whole object) and comparing them against the search criteria. Thus the operation will have a very low memory signature and will scale for huge amount of object data.

2.3. Service Layer

The main objective of the NEOBASE project is to make the data, about the neurons and synaptic connections, available for the benefit of the larger research community. The data will be freely accessible using standard internet technologies such as World Wide Web, to the interested users. In addition to providing data querying and browsing facilities to the individual scientists, we aim to accommodate a large number of third party application programs such as programs for constructing 3D representation of neocortical microcircuits, tools for analyzing neuron and synapse data based on various models etc.

The data in the NEOBASE will be accessible, to the user, via an easy to use and comprehensive web portal. Additionally, the database will also be accessible, via a web service’s layer, to the third party application tools for circuit building, visualization and simulation. The web portal will facilitate individual users to connect to and browse the microcircuit data via standard web browsers. However, the application programs will require a programmatic access to the database in order to extract required information. It is also important to mention that on the WWW, the multi dimensional data presentations of the neurons and neuron networks is limited by the facilities offered by the HTML and related technologies. For example, it will not be possible on the HTML based pages, to display the neurons as 3D objects capable of being rotated and zoomed into etc, due to various security and performance reasons. On the other hand using the web service’s layer, advance database browsing and visualization environments will be able to extract the data about neurons and synapses etc in the native format and provide a rich interactive multi-dimensional representation.

Various database related services will be implemented as part of the current project. These services will be accessible via the web portal and the web services. These services include, but are not limited to, database querying and browsing, data upload and download tools, data analysis and processing program etc.

2.4. Data Access Layer

The data, in the NEOBASE, will be accessed by the individual scientists as well as by the application programs for circuit reconstruction, visualization and simulation, among many others. The data access layer will provide functionality to interact with and query the data contained in the NEOBASE. The application programs are, however, responsible for interpreting data for their usage. The system will be accessible to web browsers as well as application programs i.e. for circuit building, visualization etc. The latter will use the web service’s layer for accessing the database.

---

Various technologies such as Java Applets allow for executing arbitrary complex code on the client machines, however these technologies are not widely used due to various security and performance reasons etc.
3. Current Status

A prototype version of the database has been implemented. The database consists of set of elementary classes corresponding to Neurons, Synapses and their different profiles. The data in the older version of the database (Markram et al., 2003) is being migrated to the new format. The figure 2 describes the fundamental classes used for representing the Neocortical Microcircuit.

![Database scheme for the NEOBASE project](image)

The following paragraph provides a very brief description of the classes as described in the above figure.

The Neuron class describes the general as well as profile specific properties of a microcircuit neuron. Similarly, the Synapse class represents the properties of a connection between two neurons. Both Neuron class and Synapse class contain data members corresponding to Attribute class and Profile classes respectively. The Attribute class describes a property of the neurons and synapses. These properties, in turn, may have values corresponding to the Value class. The Value class is extended to
StringValue, BooleanValue, FloatValue, IntegerValue and CompositeValue in order to describe different data types. Similarly, the Profile class is subclasses into MorphologyProfile, ElectrophysiologyProfile, GeneExpressionProfile (and synapse profile classes) in order to represent different neuron and synapse profiles as described in the section 1 of the current paper. The current version of the database consists of classes facilitating storage of neuron and synapse data only. The strategy adapted, in the NEOBASE project, is to validate the database design by migrating the existing neuron and synapse data stored in the previous version of the database platform (Markram et al 2003). Subsequently, the framework will be extended to include classes for genes, ion channels, receptors, neurotransmitters as well.

4. Conclusion

The neocortex subserves the most sophisticated perceptual and cognitive functions of mammals and occupies nearly 80% of the human brain. The microcircuitry of the neocortex lies at the heart of this immense computational power and deriving the blueprint of the neocortical design will be essential to fully understanding neocortical information processing. The microcircuit is immensely complex with 10’s of thousands of neurons making up a functional unit depending on the species and brain region. There are a large number of different types of neurons and each is intricately mapped onto a specific fraction of its neighbors using millions of synaptic connections, each with a characteristic anatomy and physiology. It is therefore simply impossible to fully understand the neocortical microcircuit without a highly organized database. A progressively larger number of research labs are also now generating quantitative data that could greatly accelerate the quest to reconstruct the neocortical microcircuit. It is important to mention that while most of the data in the database will consist of rats somatosensory data obtained from brain slices, the task of dissecting the microcircuit in one species, age group and brain region is an immense one and will provide the foundation for selected and meaningful comparative studies in vivo, in other species and brain regions. NEOBASE will therefore expand, iterate and converge towards progressively more accurate descriptions of the neocortical microcircuits in many species and brain regions. Finally, recent studies strongly implicate microcircuit deficits, such as deficits in specific types of interneurons, in a large number of all neurological and psychiatric disorders due to migration and differentiation abnormalities, and being able to eventually simulate neocortical microcircuit behavior may be key to predicting the impact of such deficits on neocortical function.

Acknowledgments

The NEOBASE builds on the earlier work as reported in the (Markram et al 2003). We thank Rene Brun and Fons Rademakers at European Organization for Nuclear Research (CERN) for helping us to design an object oriented schema for the microcircuit database. We also thank Maria Toledo-Rodriguez and Gilad Silberberg at Brain Mind Institute (EPFL) for their help in the project conceptualization and initiation.
References:


