

The GENESIS Simulation-based Neural Modeling Database

Jason Leigh (spiff@eecs.uic.edu), Thomas A. DeFanti
Electronic Visualization Laboratory, University of Illinois at Chicago

Brad B. Blumenthal
Interactive Computing Environments Laboratory, University of Illinois at Chicago

Robert Grossman
Laboratory for Advanced Computing, University of Illinois at Chicago

James M. Bower, David H. Bilitch
Department of Computational Neural Systems, California Institute of Technology

David Beeman
University of Colorado at Boulder

February 24, 1994

Abstract

This paper presents a multidisciplinary collaboration between neurobiologists, database-, visualization- and human/computer interaction specialists, in the design of a database for the storage of simulation models from Computational Neuroscience¹. The design consists of using the neural simulation system GENESIS, as the underlying model for the scientific database. In addition we briefly describe some of our preliminary work in using Virtual Reality interfaces to access scientific databases.

1 Introduction

The Human Brain Project (HBP) is a broadly based federal research initiative whose purpose is to encourage and support research on basic and clinical neuroscience, as well as informatics which could be used to facilitate neuroscience research [GCA⁺93, PM91]. Neuroscience is an extremely broad field involving the study of behavior at the highest level to the study of genes contained in cells of the nervous system, at the lowest level. One of the ultimate goals of the HBP initiative is to construct a database system that can record and interrelate information from every aspect of neuroscience research.

¹This work is supported by The Human Brain Project of the NIMH, grant # MH52145
Additional support includes: NSF grant # IRI-9213822, and # CDA-9303433 which is also supported by ARPA

1.1 Computational Neuroscience

The last several years have seen a dramatic increase in the number of neurobiologists building or using computer-based models as a regular part of their efforts to understand how different neural systems function [Bow92]. As experimental data continues to amass, it is increasingly clear that detailed physiological and anatomical data alone are not enough to infer how neural circuits work. Experimentalists appear to be recognizing the need for the quantitative approach to exploring the functional consequences of particular neuronal features which is provided by modeling. This combination of modeling and experimental work has led to the creation of the new discipline of computational neuroscience [EB93].

More than the use of models *per se*, computational neuroscience is most distinguished from classical neurobiology by an explicit focus on how the nervous system computes. Thus, instead of obtaining experimental information about the structure of the nervous system for its own sake, a computational approach involves collecting that information most relevant at the moment for the advancement of functional understanding. In the long run, this will undoubtedly change the organization of neuroscience. For example, neurobiology laboratories have traditionally been organized around one or two experimental techniques. Now, however, increasing numbers of laboratories are using whatever experimental technique is necessary to obtain a piece of information that a model indicates is necessary. In some cases this even involves developing new experimental technology. This results in one laboratory applying a much wider range of experimental techniques.

The driving force behind these changes is clearly the development of computer models which, in effect, demand that specific types of experimental data be obtained. It is these same models that can then provide an interpretation for the obtained data. Consequently models become an encapsulation of our most up-to-date understanding of the nervous system. This is important in the context of the human brain database, because the emphasis is then placed on the storage of our *understanding* of the nervous system, rather than the storage of mass amounts of possibly uncorrelated experimental data.

To support the interaction between experimentation and modeling Caltech has developed over the last eight years, the GENESIS neural simulation system. As a result an increasing number of simulation models are

beginning to emerge; and yet there does not exist an efficient mechanism by which the knowledge embedded in the models can be consolidated. We are now in the preliminary design stages of building a GENESIS-based neural modeling database for this purpose. This paper presents our multidisciplinary approach to the design of this database. The work is the product of a collaboration between human/computer interaction, database, and scientific visualization researchers, at the University of Illinois at Chicago, and neurobiologists at the California Institute of Technology.

2 Multidisciplinary Approach to Scientific Database Design

Database research has traditionally centered on providing a rigorous theoretical basis for the storage and manipulation of data. In designing a neuroscience database, traditional approaches would have concentrated more on the database than on the neuroscientists; typically resulting in a system that relegated them to using SQL-like interfaces to access the data. This is an unnatural interface for neuroscientists because it forces them to conform to a means of database access that is more suited to the computer than to the user. Moreover such unextended query interfaces cannot support the complexity of many neuroscience queries [BM93]. Though we do not suggest that SQL-like interfaces be abandoned, we do recommend that non-traditional interface designs be considered also. Our approach extends the bounds of traditional database design to include current research in human/computer interaction and visualization [LDSL⁺93, CLB⁺93, CSD93]. Here the focus is placed on both the need to design a system that supports the *process* of neuroscience research, as well as the need to design a rigorous database architecture[FJP90, BK90].

User studies [Pre93, BB87] are common tools used by user-interface specialists at practically every phase of a software life-cycle. It is especially important in the initial stages of the cycle, because it allows designers to clearly define the needs of the users before any design is attempted. Having understood these needs, prototypes may then be built to test design ideas. Additional user studies may then be performed on the prototypes to give the users another opportunity to respond to the design. This prototyping/user testing process may be repeated a number of times before any final decisions on the software is made.

2.1 Characterizing the Neuroscience Research Process

In working with the Division of Biology at Caltech over the last two years, we have managed to form a comprehensive understanding of the process of neurobiological research. This was achieved through many laboratory meetings with neurobiologists, and their responses to an electronic questionnaire produced using the Mosaic hypermedia system [And93].

A hypertext system is a textual database where related information is connected by links in a graph structure. An information element is a node in the structure which is usually presented as a screenful of text, and more recently, graphics and sound (hence the term hypermedia). The links that exist between a currently displayed node and others are activated by the selection of an anchor. This results in the retrieval of one or more nodes [Con87, FC89]. A hypertext document has the advantage of allowing the user to choose the level of detail of the material he/she is reading. This is typically done by following links in a depth-first fashion. The drawback of hypertext systems however is that it can cause the user to become lost in a maze of information elements [Nie90, Tra89].

Mosaic is a hypermedia system developed by the National Center for Supercomputing Applications to support the wide-area-networked access of hypertext documents that include text, graphics, animation and sound. The networking capabilities of Mosaic were particularly suited to our task because it allowed us to easily share questionnaires with other neurobiologists on the Internet while logging all their responses at one site for analysis.

For our questionnaire we produced a hypertext document that compiled what we (the computer scientists) understood, from our meetings with the neurobiologists, were the main tasks performed by them in their research. The purpose was to ensure that both parties shared a common mental model [Nor88] with respect to the characterization of the neurobiological research process.

The hypertext document provided text entry boxes for (anonymous if desired) responses by the neurobiologists. In addition a review button on the document was provided for the neurobiologists to view comments that were previously submitted by their colleagues (to encourage discussion). A copy of the questionnaire is shown in Appendix A. Since the questionnaire is a hypertext document we will only show the first hypertext page used

to explain the survey to the users.

From these questionnaires we were able to isolate a number of tasks that the neurobiologists agreed was an accurate characterization of their research process. These tasks include:

1. Experimentation - which involves preparing experiments in a laboratory that results in the recording of electrical activity from brain tissue samples. The data is recorded using a recording rig that amplifies signals from electrodes applied to the specimen. The signal may or may not be filtered before digitization and storage on harddisk. Each recording session is stored in a separate flat file.
2. Simulation- Based on the analysis of experimental data found from direct experimentation and/or surveying the available literature, a simulation is constructed. In the single cell case, this consists of building a neuron out of a number of discrete simulation compartments and tuning the properties of each of the compartments based on the experimental data. This may involve the addition and tuning of a number of different ionic channels that are known to exist in the neuron. The simulation results generated are then compared with experimental findings. This comparison is the primary means by which a model is validated.
3. Analysis of Data - Experimental or simulation data is translated to formats that can be imported into statistical analysis packages. Typically the biologist must write his/her own programs to perform this translation. Additional visualization tools may also be applied to gain greater insight into the data.
4. Journal Research - This is similar to research that goes on in most scientific disciplines. Journals and books are searched for relevant material. References are tracked to obtain further relevant information. Photocopies may be made for particularly relevant material which may include graphical as well as textual data. For example it may include graph plots of experimental data.
5. Personal Note Taking and Data Management - Personal notes and notebooks are kept to log experimental techniques and map the search space of possible research paths. Additional notes consist of diagrams, graphs, equations, simulation parameters for experimental and simulation work. Data management involves the storage of data from experiments, simulations and results of analyses, in flat files for later use.

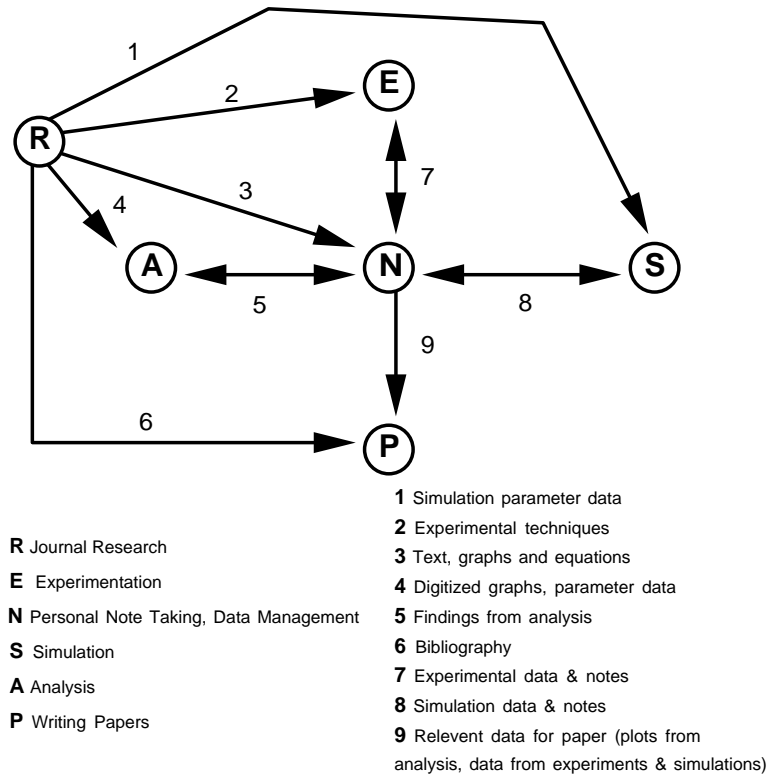


Figure 1: The dataflow between tasks typically performed in neuroscience research.

6. Publishing Results - Results are published through standard peer-reviewed mechanisms such as journals and conference proceedings. Various image conversions are performed to transfer graphs from analysis packages and drawing programs to word processors. Bibliography lists have to be compiled.

Figure 1 illustrates the relationship between the individual tasks in terms of the data flow between them. From the graph it can be seen that personal note taking and data management (node N) is a central part of all aspects of the research process. Transitions to- and from this node amounts to a translation from one data format to another. However from our surveys we have learned that these translations are performed largely by personalized data translators. That is, researchers will, for example, write their own translators to transform data gathered from experiments to forms acceptable to their data analysis programs (transition from node N to node A). Also node N represents a personal working set of data that is relevant to the current research problem and an archive of research done in the past. Some researchers will write their own tools to create a small database from experimental data (transition from node E to node N). These tools are specific to the type of data generated

from their own experiments and are usually not compatible with those of other researchers.

Also there is a high dependence on previously published data (node R) in practically every task, although transitions from node R was one of the least computer-assisted. It was identified that data from node R did not simply consist of information from previously published literature, but also included simulation models and parameters from other researchers in the field.

From understanding the neuroscience research process we were tempted to build the tools we felt would be useful to the neurobiologists. Fortunately we chose instead, to ask the neurobiologists what they *really* needed. This prevented us from intruding on the part of neuroscience research, for which the neurobiologists did not want computerized assistance.

The following is a compilation of what the neurobiologists felt would be useful to their research:

1. It would be useful to integrate the simulation system, and data from experiments with a data analysis and visualization system. That is the neurobiologists want a set of tools to convert data from their experiments and simulations to formats that can be read by their analysis and visualization tools.
2. It would be useful to maintain an electronic library of neuroscience literature from which some of the parameters used in modeling work can be systematically collected to allow easier access and comparison.
3. It would be useful to have a means of maintaining a large collection of modeling data. Multiple simulation models for perhaps the same neuron should be stored to allow comparison. Simulation parameters should be retrievable for other simulation work. Also it is important that models stored in this collection have references to the literature and information on experimental techniques from which it was derived. There should be an automatic mechanism for linking modeling data with experimental data and information found in electronic copies of neuroscience literature.
4. It would be useful to maintain a personal database of notes, diagrams, equations, experimental and modeling data.

Figure 2 shows a grouping of the tasks of Figure 1 into three main software supportable clusters that are

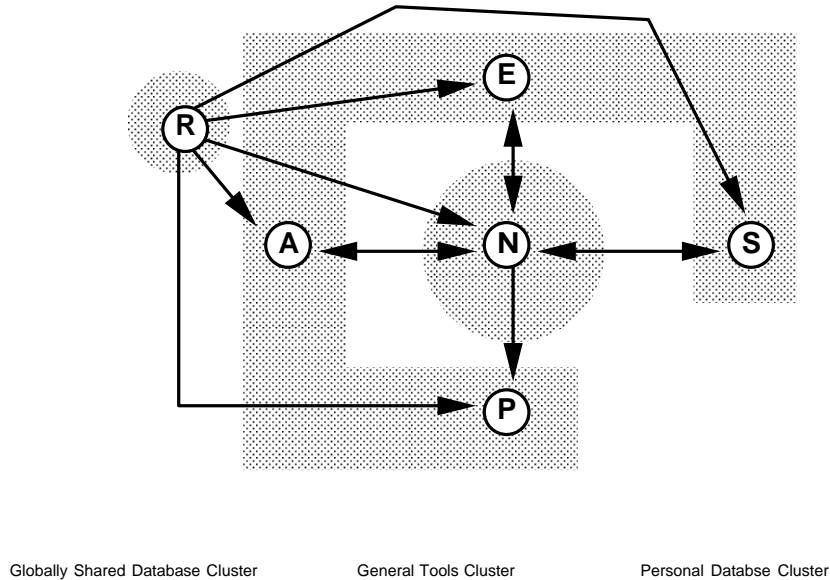


Figure 2: Grouping of tasks into three main software supportable clusters.

motivated by the set of needs described above. The first cluster (personal database cluster) consists of a database system to maintain personal research notes, and experimental and simulation data. The second cluster groups the experimentation task with the simulation and data analysis tasks (general tools cluster). The final cluster consists of a database of neuroscience literature and GENESIS simulation data (globally shared database cluster).

Other researchers[BG91, WI93] are already actively working on developing generic databases for experimental data. Our interest is in the design and construction of the globally shared database cluster (that is, the GENESIS Simulation-based Neural Modeling Database). In parallel with this effort, we intend to continue our work on developing advanced visualization tools to view neuroscience data[CLB⁺93, LDSL⁺93, LDG⁺94]. The next section discusses our proposed design of the simulation-based neural modeling database.

3 GENESIS Simulation Based Neural Modeling Database Design

The GENESIS Simulation-based Neural Modeling Database will consist of three main components. Firstly a hypertext/hypermedia database will contain electronic copies of neuroscience literature and references. These references will be maintained in formats that are readily usable in current popular word processors.

Secondly a database of simulation models and data will be maintained by an extended version of the GEN-

ESIS simulator. The attractiveness of using a simulator as a database is that queries that require information regarding the function of a neuronal system can be answered by executing a simulation—based on what GENESIS understands about the particular neuronal system in question it can “synthesize” data that does not currently exist in the database. By traditional database means this would require instantiating every possible outcome of a simulation model.

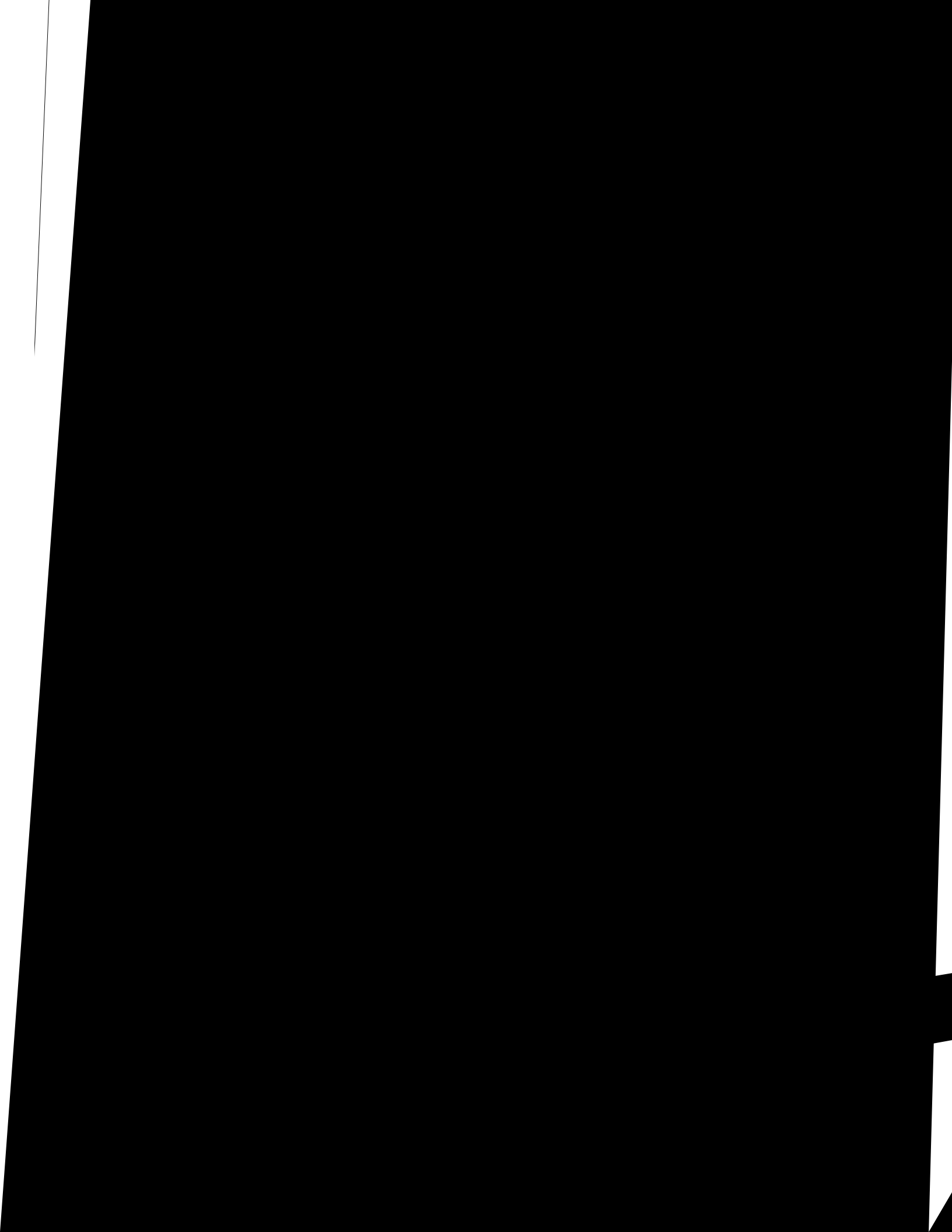
As it is important for all modeling data be firmly based on experimental sources, a third database will maintain small snapshots of relevant experimental data.

Each of these three components will be discussed individually in the next section. The relationship between these three databases are shown in Figure 3. Data from each database will have relevant links to information in the other two databases.

Finally there will be a number of user-interfaces to the simulation-based modeling database. This is intended to support the needs of our neurobiologists as well as others in the framework of the human brain project. Higher level interfaces will be provided for casual users of the system to provide browsing while lower level interfaces will be provided for experienced users who wish to quickly access specific pieces of information. Extensions to the interfaces as a whole will include a means of translating data retrieved from queries to formats that can be read by external data analysis and visualization programs.

3.1 GENESIS

GENESIS (for GEneral NEural SIMulation System) has been developed at Caltech as a research tool to provide a standard and flexible means of constructing realistic simulations of biological neural systems. The system was specifically designed to allow the construction of biological simulations at many different levels, from sub-cellular components, to whole cells to networks of cells [BH91]. The ultimate objective was to provide a simulation platform that could support simulations of the nervous system at any level of detail and complexity.



amenable to this approach because they typically consist of discrete components interacting in quite stereotyped ways and because the different simulations tend to use similar neural components, display routines, numerical integration routines, etc.

User Extendible

This object-based approach is central to the generality and flexibility of the GENESIS system [BH91]. For example, this modularity means that it is possible to quickly construct a new simulation or to modify an existing simulation by inserting different simulation objects from the existing library of standard simulation components. In this way, individual modules or linked assemblies of modules (such as compartments with channels, entire cells, or networks of cells) may be easily replicated. In addition, and importantly, the object-based nature of GENESIS also allows for users to flexibly extend the capacities of the system by constructing new simulation components not yet included in the object libraries. In this way, the simulation objects available for use within GENESIS continues to grow as the system is used. This growth is reflected in the size and complexity of the libraries of GENESIS objects and simulations that currently exist.

GENESIS Libraries

Current GENESIS libraries are quite extensive and have been contributed to by researchers from all over the world. As such they represent a continuously expanding detailed description of the nervous system. GENESIS libraries can be divided into two types, those that contain whole simulations of cells or networks, and those that contain the components from which these simulations are constructed. These two aspects will be discussed separately.

Simulation Object Libraries

Simulation object libraries contain the building blocks from which many different simulations can be constructed. These are essentially class libraries from which objects are instantiated, and from which the database schema is composed. For example there are classes to define active and passive compartments; both voltage and synaptically activated channels; individual gates; dendro-dendritic channels; and synapses of several types including Hebbian synapses. In addition, there are classes for computing intracellular ionic concentrations from

channel currents; classes to allow ligand gating of ion channels, and axons to provide connections to other neurons, with specified propagation delays and synaptic weights. There are also a number of device classes to provide various types of input to the simulation (pulse and spike generators, voltage clamp circuitry, etc.) or measurements (peristimulus and interspike interval measurements, spike frequency measurements, etc). As in Object-Oriented programming, each class contains members that define the state of an object of that class, as well as member functions that define the (mathematical) behavior of the object.

Simulation Libraries

In addition to the GENESIS libraries which contain the components from which simulations are constructed, there are also libraries of completed simulations and models constructed from these components. For example, detailed single cell models alone currently include: cerebral cortical pyramidal cells; Hypocampal pyramidal cells [DSB93]; cerebellar Purkinje cells [DSB92, Bow92]; and mitral, granule, and tufted cells from the olfactory bulb [BB93].

As already mentioned, GENESIS is designed not only to simulate single neurons, but also networks of neurons. While the majority of the work in computational neuroscience to date has focused on single cell models, larger network models are starting to be built. Simulation libraries of circuits include the olfactory [WB92] and visual cerebral cortices [WB91], the olfactory bulb [BB93], the inferior olive, and the central pattern generator of tritonia [RBM89]. As with brain component libraries, these simulations contain detailed information about the organization of different brain regions.

Visualizing GENESIS Output

XODUS (for X-based Output and Display Utility for Simulators) provides users with a set of graphical objects (widgets) for outputting results obtained during simulations, as well as inputting new parameters between simulations. Essentially this means that users do not have to write low-level X window code to produce a graphical interface to their simulation. High-level widgets already exist to provide the familiar buttons and potentiometers found in most graphical interfaces, as well as tools that are neuro-modeling specific (for example there is a widget that draws three-dimensional representations of neurons).

As XODUS runs under GENESIS, its ability to offer high speed visualization is severely limited. Consequently external visualization tools (GENESIS Visualizer and GENESIS Listener) [LDSL⁺93] have been developed to take advantage of the display capabilities of currently available high-performance graphics workstations.

GENESIS Visualizer allows users to load pre-computed simulation data for high-speed animated playback and recording, either on film or video. This allows the viewer to gain a holistic understanding of the dynamical process being viewed that ordinarily could not be gained by watching static images [LDSL⁺93].

GENESIS Listener is a program executed concurrently with a GENESIS simulation to allow viewing of data as it is being generated by the simulation. This was developed primarily to provide a minimal graphical interface to a version of GENESIS running on the Intel Touchstone Delta supercomputer. One of the problems with using supercomputers for simulation work is that processing is done in batch mode. Most supercomputers do not yet support an X windowing environment for interaction. There is no opportunity to interact with the simulation during execution. GENESIS Listener provides a graphical front-end which communicates with the simulation by a network socket connection. The user is therefore able to view and manipulate the data as it is being computed by the simulation. At the same time the user can provide new input parameters to steer the simulation in progress.

We have also built a version of the GENESIS Listener to provide an interface between GENESIS and the CAVE virtual reality system (Figure 5) [LDSL⁺93, CLB⁺93]. The CAVE (for CAVE Automatic Virtual Environment Audio Visual Experience) is a room composed of four, ten foot projection screens (left, right, front and floor screens) on which alternating stereoscopic computer-generated images are displayed. The user dons a pair of shutter glasses that are synchronized with the projections. The visual result is that the projections appear to take on a three-dimensional form. Users can interact with virtual objects in this environment by using a six dimensional pointing device. Work is still in progress on this.

3.2 GENESIS Database Extensions

Many problems associated with designing scientific databases have been identified by numerous researchers in the past [WI93, Lan91, Fre91, FJP90, GQ93, GLQ93, BG91, BM93, BK90, Kim90]. Our approach takes a

Figure 5: The CAVE is a high-end virtual reality system. This diagram shows what a virtual neuron would look like projected in the CAVE and viewed with the stereographics glasses.

simulation system that is already familiar to a large number of computational neuroscientists, and extends it to provide database capabilities. This has similarities to the work done by Ioannidis[IL89]. Here we describe how our use of a simulation system can help mitigate many traditional problems facing scientific databases. We do not claim that this approach solves all the problems associated with scientific database design; our motivation for using this model, as we had mentioned earlier, is based largely on the needs of the neurobiologists and their familiarity with GENESIS.

3.2.1 Choosing a Database Architecture: Relational vs Object-Oriented vs Simulation-based

Object-oriented databases are generally preferred over relational databases for scientific data storage because of their ability to naturally represent the hierarchical structure inherent in scientific data [BM93, Kim90, BG91]. They are also favored for their ability to encapsulate behavior that is not readily representable in a relational database. Relational databases on the otherhand are favored for having an established mathematical model (relational calculus)[Kim90, BM93].

As described earlier, GENESIS currently amasses a great deal of information regarding the nervous system. To encode this into a traditional relational or object-oriented database would require an enormous effort. Not only will discrete values have to be encoded, but as mentioned earlier, simulation characteristics (meta-data) must also be modeled. The resulting effort would no doubt amount to turning a relational or object-oriented database into a GENESIS simulator.

Another problem with using a generic relational or object-oriented database is that neither usually provide powerful enough query languages to express scientific queries [BM93, Fre91]. As a result additional programs have to be written that access the database through an application programmer's interface[BM93, Fre91]. GENESIS on the otherhand provides a neural modeling-specific language that can be extended to provide the functionality of a database query language. This will allow neurobiologists to express queries in terms they are already familiar.

3.2.2 Quality Assessment of the Data

One of the most difficult aspects of constructing a database is assuring that all of the data entered is accurate [FJP90]. Usually, this requires a moderator capable of certifying the quality of the data. The more inaccurate the stored data, the less useful the database becomes (the analogy is that a 5% error in a relation joined over 15 relations can result in an outcome being less than 50% accurate) [Fre91]. In the current case the data from which the database will be derived is contained in numerical simulations already shown to be capable of replicating specific types of brain activity. While determining the absolute accuracy of a particular model or model parameter is a complex process [BB93], the user of the database at least knows that values are within an appropriate range for the specific behavior modeled. In this way the simulations themselves represent an internal check on the accuracy of the data in the database.

It is often the case that data obtained by different experimentalists conflict or appears to conflict. Not infrequently these conflicts are difficult to resolve at first glance. Modeling often not only highlights these conflicts, but can also potentially provide an opportunity for their resolution. Thus, not only do models point out data conflicts, they can also help to resolve the conflict, thus serving again, as a check on the veracity of the data in the database.

3.2.3 Functional Significance of the Data

Another issue related to the quality of stored data involves the question of what data is most relevant and therefore useful to our knowledge of a particular system at a particular time. In other words, just because data can be collected does not necessarily mean that it will help expand our current understanding of a particular neural system [Fre91, FJP90]. However, if the data has been included in a functioning biological model and can be demonstrated as necessary to produce the desired output [BB93], then some basis for relevance can be established. Further by exploring a particular model it is possible to determine exactly how the information is relevant to that model's output.

Ultimately we assume the objective of the Human Brain Project was not simply to store massive amounts of

data, but to contribute to our understanding of its significance for human brain function. However, with large and growing databases of the most common type, it is not at all clear how the data is eventually assembled to create some functional understanding. In the current case, however, the basis for exploring the functional implications of the data are built into the design and construction of the database itself. The models that serve as the point of entry for the data are also tools that can be used to understand its significance. As the models become more sophisticated, so does the representation of the data. As the models become more capable, they extend our ability to explore the functional significance of nervous system structure and organization. Thus, there is a direct link between the ultimate objective of acquiring the data and the data acquisition process itself.

3.2.4 Handling Massive Amounts of Data

Simulations represent the most compact form of data possible. For example, in principle, a simulation capable of replicating all the features of the biological system it mimics could reconstruct whatever dataset is of interest from first principles. In this case, raw data would not need to be stored at all. While we DO NOT anticipate that this will happen any time soon, it does illustrate the fact that models can be, in effect, a very compact means of representing data. For example, a correctly parameterized Hodgkin Huxley model in principle can contain all the information necessary to reconstruct voltage and current clamp records for a single population of channels.

3.3 Hypermedia Database

This database will implement a hypermedia system to store references from the individual components of GENESIS data to electronic versions of referenced papers. With this capability, the user can follow a hypertext link to the relevant references from any piece of data extracted from the GENESIS database. Similarly by browsing a hypertext document a user can retrieve data from the GENESIS database. A hypermedia system is provided rather than a straight-forward hypertext system because we anticipate the need to store graphical and possibly audio information.

3.4 The Experimental Data Store

The parameters that go into a successful simulation model in GENESIS are largely derived from experimental data. Neuroscientists who build realistic models tune those models based on comparisons of model output with experimental results. Often this is done by “eyeballing” the data. However, we have recently developed techniques that allow a quantitative comparison for detailed parameter searches if the experimental data is available in digital form [BB93]. We believe that such quantitative comparisons will be used increasingly in neural modeling [BC92]. This means it will be possible to associate each simulation with the specific experimental data the modeler has found most useful in the model tuning and testing process. We propose to link this experimental data to the models in the database. Note, however, that we do not intend to build a massive data store of raw experimental data but only to store the experimental data that is most relevant to the models in the GENESIS database. This means that the stored data always has a context.

The underlying model of the experimental database is a persistent object store called PTool.

3.5 PTool

PTool is a persistent object-oriented storage tool that is currently under development at the Laboratory for Advanced Computing at the University of Illinois at Chicago [GQ93, GLQ93, BG91]. In collaboration with Argonne National Laboratory, Ptool has been used to store the terabytes of data generated by particle collisions in collider experiments at the Collider Detector Facility at Fermi Lab [GQ93]. It has been used to store approximately a million trajectory segments representing the possible paths of nonlinear systems describing aerospace and robotic systems [GLQ93], and recently, it has been used to store all the tables in the 1993 federal budget.

Ptool may be viewed as a persistent object store with objects arranged over a hierarchical storage system. Objects are gathered together into physical volumes. One or more volumes are wrapped into bitfiles, which are managed by the hierarchical storage system. To improve efficiency, volumes are divided into physical segments, which are cached and migrated as appropriate. Volumes are used to transport data from the object level of the system to the file level. The system is divided into a number of components. Each component is implemented as a tool. The tools are arranged into layers, which transport data from one level to an adjacent one (Figure 6).

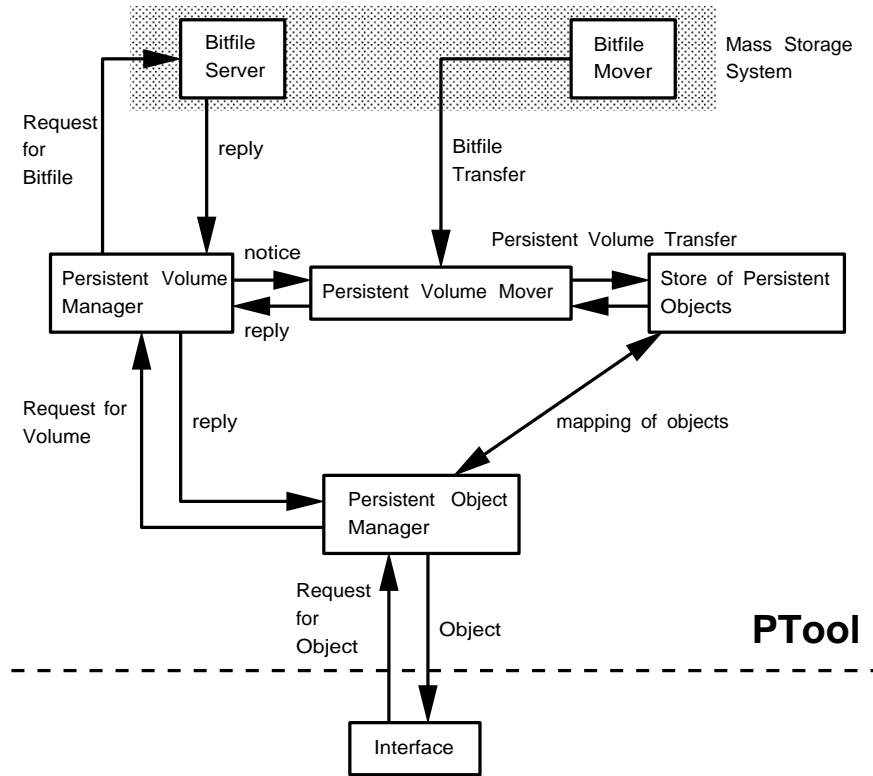


Figure 6: The PTool architecture.

The top level of Ptool consists of tools to store and retrieve objects. Currently this is done through the use of C or C++ calls to the Persistent Object Manager, described below.

The second level consists of tools which provide basic support services for objects as needed, for example maintaining the consistency of concurrent transactions.

The third level consists of a Persistent Object Manager to create, access, and store persistent objects. Collections of various sorts are provided by using the appropriate class libraries. Collections themselves may be transient or persistent; if persistent, they are handled by the Persistent Object Manager as any other persistent object. Persistent objects are objects that exist independent of the processes that create them.

The fourth level consists of tools to move physical collections of persistent objects between the file system and the Persistent Object Manager. This consists of the Persistent Volume Manager and the Persistent Volume Mover. The Persistent Store (or depot) is divided into physical regions called volumes. If the Persistent Object Manager determines that the object requested is not in a volume currently available at the depot, it sends a

request for the volume to the Persistent Volume Manager. The Persistent Volume Manager determines the bitfile containing the volume and sends a request for the bit file to the Bitfile Server, part of the storage system. The bitfile is returned by the Bitfile Mover, also part of the storage system, to the Persistent Volume Mover, which extracts the volume from the bitfile, loads the persistent volume into the local cache for the depot, and sends a reply to the Persistent Object Manager indicating that the volume is loaded. For efficiency, volumes are divided into physical segments and segments are migrated, rather than entire volumes. Segments from a variety of volumes are available at a depot at the same time.

The bottom level consists of a hierarchical storage system to access and store bitfiles, containing volumes of persistent objects.

3.6 The User-Interface

We will provide multiple user interfaces tailored to the anticipated different needs of the users. A low-level query interface will allow users to perform complex searches on modeling data. We anticipate that this is the interface that will be most used by modelers and researchers. In addition, a high-level query interface will allow browsing of neuroscience papers and modeling data. This is useful for first-time users of the system.

As we mentioned earlier, we are interested in exploring other interface possibilities beyond the traditional text-based query languages. For example, we have recently constructed a low-cost virtual reality system (Figure 7) that we believe is more accessible (affordable) to neurobiologists. This system consists of a single graphics workstation with the same stereoscopic shutter glasses used in the CAVE, and a six dimensional pointing device for interaction [LDA⁺93, LDG⁺94]. With this system we were able to visualize: a simulation of the electrical activity of a weakly-electric fish; a network simulation of the piriform cortex; and a simulation of a cerebellar Purkinje cell[LDA⁺93, LDG⁺94]. These virtual reality demonstrations are currently used as prototypes for our research into the use of virtual reality techniques to explore scientific databases [JF94, LDG⁺94]. For example, in the Purkinje cell prototype (Figure 8) we interface our virtual reality system with PTool. PTool maintains all the data from a Purkinje cell simulation while the virtual reality system allows the user to use a virtual electrode to probe various parts of a three-dimensional representation of the cell, to take readings. This is

Figure 7: A low-cost/high-performance virtual reality system for neuroscientists. The 6D mouse is used as a pointing device in 3-space. The Crimson Reality Engine was replaced by a lower-end Indigo in the earlier version of the system.

equivalent to making queries in a traditional database using a query language except the interface is, we believe, more natural to neurobiologists because it presents them with familiar tools to suggest the kinds of database operations possible. We are still in the process of evaluating the relative benefits of using such a system over traditional query languages for database access.

4 Future Work

Our approach to database design has involved a diverse group of experts in computer science and neurobiology. This collaboration has produced a rich source of ideas and techniques that ordinarily would have remained independent of each other.

Having thoroughly researched the needs of our neurobiologists, and outlined a possible design to meet these needs our next step will be to develop interface prototypes which will be used to evaluate design ideas. Our visualization group will continue to develop visualization tools for viewing neuroscience data, as well as novel

Figure 8: Purkinje cell cycling through data stored on a PTool data store file. The virtual electrode is used to query individual components of the neuron for its membrane potential.

interfaces that include the use of virtual reality to explore scientific databases. Although it has been suggested in database circles that the ultimate interfaces to databases should be generic and domain independent[BK90], we believe the area of user-interface design in scientific databases is still in its infancy. As such there is still a great deal that can be learned from developing domain-specific systems such as this one.

References

- [And93] M. Andreessen. Ncsa mosaic technical summary. Technical report, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, May 1993.
- [BB87] R. M. Baecker and W. A. S. Buxton, editors. *Readings in Human-Computer Interaction*. Morgan Kaufmann Publishers Inc., 1987.
- [BB93] U. Bhalla and J.M. Bower. Exploring parameter space in detailed single neuron models: Simulations of the mitral and granule cells of the olfactory bulb,. *Jour. of Neurophysiol.*, 1993.
- [BC92] J.M. Bower and Koch. C. Experimentalists and modelers: Can we all just get along? *Trends in Neurosci.*, 15:458–461, 1992.
- [BG91] A. Baden and R. Grossman. Database computing in high energy physics. In Y. Watase and F. Abe, editors, *Computing in High-Energy Physics 1991*, pages 59–66, Tokyo, 1991. Universal Academy Press, Inc.
- [BH91] J.M. Bower and J. Hale. Exploring neuronal circuits on graphics workstations. *Scientific Computing and Automation*, pages 35–45, March 1991.
- [BK90] F. Bancilhon and W. Kim. Object-oriented database systems: In transition. *SIGMOD RECORD*, 19(4):49–53, December 1990.
- [BM93] E. Bertino and L. Martino. *Object-Oriented Database Systems*. International Computer Science Series. Addison-Wesley, 1993.
- [Bow92] J. M. Bower. Modeling the nervous system. *Trends Neurosci.*, 15:411–412, 1992.

- [CLB⁺93] C. Cruz-Neira, J. Leigh, C. Barnes, S. M. Cohen, S. Das, R. Engelmann, R. Hudson, M. Papka, T. Roy, L. Siegel, C. A. Vasilakis, T. A. DeFanti, and D. J. Sandin. Scientists in wonderland: A report on visualization applications in the cave virtual reality environment. In *Proceedings IEEE 1993 Symposium on Research Frontiers in Virtual Reality*, pages 59–65. IEEE Computer Society Press, October 1993.
- [Con87] J Conklin. Hypertext: an introduction and survey. *IEEE Computer*, 20(9):17–41, September 1987.
- [CSD93] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In J. T. Kajiya, editor, *Computer Graphics (SIGGRAPH '93 Proceedings)*, volume 27, pages 135–142, August 1993.
- [DSB92] E. De Schutter and J. M. Bower. Purkinje neuron simulation on the intel touchstone delta with genesis. In T. Mihaly and P. Messina, editors, *Proceedings of the Grand Challenge Computing Fair*, pages 268–279, Caltech, Pasadena, CA, 1992. CCSF publishing.
- [DSB93] E. De Schutter and J.M. Bower. Sensitivity of synaptic plasticity to the ca2+ permeability of nmda-channels: A model of long term potentiation in hippocampal neurons. *Neural Computation*, 1993.
- [EB93] F. H. Eeckman and J. M. Bower, editors. *Computation and Neural Systems*. Kluwer Academic Publishers, Boston, 1993.
- [FC89] M. Frisse and S. Cousins. Information retrieval from hypertext: update on the dynamic medical handbook project. In *Proceedings of the Hypertext '89 Conference*, pages 199–212, Pittsburg, Penn., November 1989.
- [FJP90] J.C. French, A. K. Jones, and J. L. Pfaltz. Summary of the final report of the nsf workshop on scientific database management. *SIGMOD RECORD*, 19(4):32–40, December 1990.
- [Fre91] K. A. Frenkel. The human genome project and informatics. *Communications of the ACM*, 34(11):41–51, November 1991.

- [GCA⁺93] F. K. Goodwin, M. E. Clutter, D. F. Alexander, J. Vaitukaitis, S. F. Zornetaer, R. A. Millstein, G. D. Cohen, J. B. Snow Jr., D. A. B. Lindberg, and F. M. Sulzman. The human brain project: Phase i feasibility studies, program announcement. PA-93-068, NIMH, NIDA, NSF, NIA, NICHD, NIDCDM NCM NLM, ONR, NASA, April 1993.
- [GLQ93] R. Grossman, D. Lifka, and X. Qin. An object manager utilizing hierarchical storage. In *Twelfth Symposium on Mass Storage Systems*. IEEE, IEEE Press, 1993.
- [GQ93] R. Grossman and X. Qin. Ptool: A software tool for working with persistent data. Technical Report 93-5, Laboratory for Advanced Computing, University of Illinois at Chicago, 1993.
- [IL89] Y. Ioannidis and M. Livny. Moose: Modeling objects in a simulation environment. *Information Process*, pages 821–826, August 1989.
- [JF94] A. Johnson and F. Fotouhi. S.a.n.d.b.o.x. an interface to scientific data based on experimentation. In *Proceedings of the Fifth Eurographics Workshop on Visualization and Scientific Computing (submitted for review)*. Eurographics, 1994.
- [Kim90] W. Kim. Object-oriented approach to managing statistical and scientific databases. In *Proceedings of the Fifth International Conference on Statistical and Scientific Database Management*, Charlotte, N.C., April 1990. V SSDBM.
- [Lan91] E. S. Lander. Mapping and interpreting biological information. *Communications of the ACM*, 34(11):33–39, November 1991.
- [LDA⁺93] J. Leigh, T. A. DeFanti, C. Assad, B. Rasnow, A. Protopappas, E. De Schutter, and J. M. Bower. Neural simulation software demonstration exhibit. Neuroscience 93 Conference, November 1993.
- [LDG⁺94] J. Leigh, T. A. DeFanti, R. Grossman, E. De Schutter, A. Protopappas, C Assad, and J. M. Bower. Virtual reality in computational neuroscience. In *Applications of Virtual Reality (to appear)*, Liverpool, England, 1994. British Computer Society Display Group.

- [LDSL⁺93] J. Leigh, E. De Schutter, M. Lee, U. S. Bhalla, J. M. Bower, and T. A. DeFanti. Realistic modeling of brain structures with remote interaction between simulations of an inferior olivary neuron and a cerebellar Purkinje cell. In *Proceedings of the SCS Simulations Multiconference*, Arlington, VA, March 1993.
- [Nie90] J. Nielson. The art of navigating through hypertext. *Communications of the ACM*, 33(3):296–310, 1990.
- [Nor88] D. A. Norman. *The Design of Everyday Things*. Doubleday, 1988.
- [PM91] C. M. Pechura and J. B. Ed. Martin. *Mapping the Brain and its Functions*. National Academy Press, Institute of Medicine, Washington, D. C., 1991.
- [Pre93] J. Preece. *A Guide to Usability: Human Factors in Computing*. Addison Wesley, 1993.
- [RBM89] S. Ryckebusch, J.M. Bower, and C. Mead. Modeling small oscillating biological networks in vlsi. In D. Touretzky, editor, *Advances in Neural information processing systems*, pages 384–393, San Mateo, CA., 1989. Morgan Kaufmann.
- [Seg89] I. Segev. *Methods in Neuronal Modeling: From Synapses to Networks*. MIT Press, Cambridge, MA., 1989. ed. C. Koch and I. Segev.
- [Tra89] M. Travers. A visual representation for knowledge structures. In *Proceedings of the Hypertext '89 Conference*, pages 147–158, Pittsburgh, Penn., November 1989.
- [WB91] M.A. Wilson and J.M Bower. A computer simulation of oscillatory behavior in primary visual cerebral cortex. *Neural Computation*, 3:498–509, 1991.
- [WB92] M. Wilson and J. M. Bower. Cortical oscillations and temporal interactions in a computer simulation of piriform cortex. *J. Neurophysiol.*, 67:981–995, 1992.
- [WI93] J. L. Wiener and Y. Ioannidis. A moose and a fox can aid scientists with data management problems. Technical Report 1182, University of Wisconsin-Madison, October 1993.

A The Mosaic Hypertext Questionnaire

Characterization of Tasks in Neuroscience Research

This list consists of my impressions of tasks that you perform in your research. The tasks are not listed in any specific order and depending on what your specific research might be, it may include all or some of the items listed below in a number of varying permutations. I am hoping that through your feedback I will be able to understand what some of those permutations might be.

The purpose of this list is two-fold. Firstly it is to verify that my impressions of what you do is correct. Secondly it is to give you an opportunity to CONFIDENTIALLY respond to my observations. This means NO ONE will be able to read the responses except the surveyor. And even then the identity of the person who wrote the response can be kept completely anonymous if desired. Please realize that you do not have to answer every question in the survey- only those parts that interest you. In particular I am interested in getting responses about the diagram below.

Since the GENESIS database will likely include some form of hypertext interface, I thought it might be a good idea to get you accustomed to hypertext by using Xmosaic (this program) rather than gopher to present this information. You will notice that some text items below are underlined. These underlined items can be selected (by pointing and clicking with the left mouse button) to display more detailed information about the item. Since Xmosaic is heavily networked based, selecting any item will cause it to connect to a remote computer to access the required information for display. In this case the information you are reading is all stored at the University of Illinois at Chicago. This is just to forewarn you that if Xmosaic takes a little while to respond, it is likely to be because of heavy network traffic.

- Experiments
 - lab preparations (depends greatly on nature of experiment).
 - record data on recording rig; pass recorded data through filters before storing on hard disk.
(Filter is not always used)
- Analysis of recorded data
 - use matlab, excel, other visualization tools.
 - make lots and lots of plots. Stick them to the wall or organize them on screen.
 - use ruler to measure data from graphs printed in journals.
 - compare recorded data to other results in journals.
- Genesis simulations
 - construct simulation model.
 - tweak parameters using published data.
 - explore parameter space to gain an understanding of the system and stimulate new ideas to do experiments on.
- Refer to papers, books, journals for ideas & data
 - gather parameter data for simulation.
 - gather data on experimental technique.
 - gather data on experimental results.
 - gather data on morphology.
- Record personal notes

- notes on experimental procedures.
 - notes on paths followed so far.
 - collect figures & equations from various papers.
 - record channel parameters so that comparisons can be made when decisions are needed on the best one to use.
- Write research paper
 - collect bibliography from previous written papers and any new references.
 - transfer equations, graphs and diagrams from analysis packages and/or GENESIS to paper.
 - Management of all this raw data gathered from experiments
 - stored on optical disk or tape.

The diagram below shows the tasks listed above and the data that flows between them. This graph does NOT show the order in which tasks are done. The numbers in the graph are intended to label data that is being transferred (electronically, manually, or mentally) between tasks; they may be used to augment your explanations in your survey responses.

Give me an idea of how you combine these individual tasks in your research (that is, what is the order in which you perform these tasks).

What do you think about this hypertext survey?

THANK YOU IMMENSELY FOR YOUR TIME!

