BB – UPM

December 2007

Introduction and Project Scope

The Universidad Politécnica de Madrid participates in the BlueBrain Project (BB-UPM) performing different research activities. This document presents a brief introduction of the overall project followed by the research lines in which UPM is involved. This document also includes a general perspective of the participating institutions and the subproject objectives.

BB - UPM

INTRODUCTION AND PROJECT SCOPE

INTRODUCTION

The Blue Brain project is the first comprehensive attempt to reverse-engineer the mammalian brain, in order to understand brain function and dysfunction through detailed simulations [1].

In July 2005, EPFL and IBM announced an exciting new research initiative - a project to create a biologically accurate, functional model of the brain using IBM's Blue Gene supercomputer. Analogous in scope to the Genome Project, the Blue Brain will provide a huge leap in our understanding of brain function and dysfunction and help us explore solutions to intractable problems in mental health and neurological disease.

At the end of 2006, the Blue Brain project had created a model of the basic functional unit of the brain, the neocortical column. At the push of a button, the model could reconstruct biologically accurate neurons based on detailed experimental data, and automatically connect them in a biological manner, a task that involves positioning around 30 million synapses in precise 3D locations.

In November, 2007, the Blue Brain project reached an important milestone and the conclusion of its first Phase, with the announcement of an entirely new data-driven process for creating, validating, and researching the neocortical column.

UPM COLLABORATION

BBP project defines four levels of engagement in the project, these levels are:

- Originating laboratories (Tier0).
- Collaborating laboratories (Tier1).
- Participating laboratories (Tier2).
- Affiliated laboratories (Tier3).

Tier0 participants is reserved for EPFL only and the rest of the tiers represent the different levels of engagements, which have a direct influence in the licensing restrictions, funding and access to internal information.

The Universidad Politécnica de Madrid (UPM) is one of the participating institutions in the BlueBrain Project (BBP). UPM is deeply involved in the objectives and visions of the BBP being engaged as one of the collaboration institutions close to the project core. The UPM aims to be engaged in the Tier1 of the BBP.

The participation of the UPM in the BBP is performed under the name BB-UPM. BB-UPM organizes the BlueBrain-Project collaboration activities directly coordinated by the UPM. These activities include different tasks in several research lines. Some of the research lines which are close-linked to neurological experimentation are performed in collaboration with Instituto Cajal (CSIC) while the research line on informatics is achieved with the participation of different computer science partners.

BLUEBRAIN PROJECT – PHASE I

This section includes a brief description of the context, approach and achievements of the Phase I of BBP. This Phase I has been completed by the end of 2007 and it is the starting point of the next phase in which UPM is mainly involved. This section has been extracted directly form [1].

Context

The cerebral cortex, the convoluted "grey matter" that makes up 80% of the human brain, is responsible for our ability to remember, think, reflect, empathize, communicate, adapt to new situations and plan for the future. The cortex first appeared in mammals, and it has a fundamentally simple repetitive structure that is the same across all mammalian species.

The brain is populated with billions of neurons, each connected to thousands of its neighbours by dendrites and axons, a kind of biological "wiring". The brain processes information by sending electrical signals from neuron to neuron along these wires. In the cortex, neurons are organized into basic functional units, cylindrical volumes 0.5 mm wide by 2 mm high, each containing about 10,000 neurons that are connected in an intricate but consistent way. These units operate much like microcircuits in a computer. This microcircuit, known as the neocortical column (NCC), is repeated millions of times across the cortex. The difference between the brain of a mouse and the brain of a human is basically just volume - humans have many more neocortical columns and thus neurons than mice.

This structure lends itself to a systematic modelling approach. And indeed, the first step of the Blue Brain project is to re-create this fundamental microcircuit, down to the level of biologically accurate individual neurons. The microcircuit can then be used in simulations.

Building the Microcircuits

Modelling Neurons

Neurons are not all alike - they come in a variety of complex shapes. The precise shape and structure of a neuron influences its electrical properties and connectivity with other neurons. A neuron's electrical properties are determined to a large extent by a variety of ion channels distributed in varying densities throughout the cell's membrane. Scientists have been collecting data on neuron morphology and electrical

behaviour of the juvenile rat in the laboratory for many years, and this data is used as the basis for a model that is run on the Blue Gene to recreate each of the 10,000 neurons in the NCC.

Modelling Connections

To model the neocortical column, it is essential to understand the composition, density and distribution of the numerous cortical cell types. Each class of cells is present in specific layers of the column. The precise density of each cell type and the volume of the space it occupies provide essential information for cell positioning and constructing the foundation of the cortical circuit. Each neuron is connected to thousands of its neighbours at points where their dendrites or axons touch, known as synapses. In a column with 10,000 neurons, this translates into trillions of possible connections. The Blue Gene is used in this extremely computationally intensive calculation to fix the synapse locations, "jiggling" individual neurons in 3D space to find the optimal connection scenario.

Modelling the column

The result of all these calculations is a re-creation, at the cellular level, of the neocortical column, the basic microcircuit of the brain. In this case, it's the cortical column of a juvenile rat. This is the only biologically accurate replica to date of the NCC - the neurons are biologically realistic and their connectivity is optimized. This would be impossible without the huge computational capacity of the Blue Gene. A model of the NCC was completed at the end of 2006.

In November, 2007, The Blue Brain Project officially announced the conclusion of Phase I of the project, with three specific achievements:

1. A new modelling framework for automatic, on-demand construction of neural circuits built from biological data.
2. A new simulation and calibration process that automatically and systematically analyzes the biological accuracy and consistency of each revision of the model.
3. The first cellular-level neocortical column model built entirely from biological data that can now serve as a key tool for simulation-based research

Simulating the microcircuit

Once the microcircuit is built, the exciting work of making the circuit function can begin. All the 8192 processors of the Blue Gene are pressed into service, in a massively parallel computation solving the complex mathematical equations that govern the electrical activity in each neuron when a stimulus is applied. As the electrical impulse travels from neuron to neuron, the results are communicated via inter-processor communication (MPI). Currently, the time required to simulate the circuit is about two orders of magnitude larger than the actual biological time simulated. The Blue Brain team is working to streamline the computation so that the circuit can function in real time - meaning that 1 second of activity can be modelled in one second.

Interpreting the results

Running the Blue Brain simulation generates huge amounts of data. Analyses of individual neurons must be repeated thousands of times. And analyses dealing with the network activity must deal with data that easily reaches hundreds of gigabytes per second of simulation. Using massively parallel computers the data can be analyzed where it is created (server-side analysis for experimental data, online analysis during simulation).

Given the geometric complexity of the column, a visual exploration of the circuit is an important part of the analysis. Mapping the simulation data onto the morphology is invaluable for an immediate verification of single cell activity as well as network phenomena. Architects at EPFL have worked with the Blue Brain developers to design a visualization interface that translates the Blue Gene data into a 3D visual representation of the column. A different supercomputer is used for this computationally intensive task. The visualization of the neurons' shapes is a challenging task given the fact that a column of 10,000 neurons rendered in high quality mesh (see picture) accounts for essentially 1 billion triangles for which about 100GB of management data is required. Simulation data with a resolution of electrical compartments for each neuron accounts for another 150GB. As the electrical impulse travels through the column, neurons light up and change colour as they become electrically active.

A visual interface makes it possible to quickly identify areas of interest that can then be studied more extensively using further simulations. A visual representation can also be used to compare the simulation results with experiments that show electrical activity in the brain. This calibration - comparing the functioning of the Blue Brain circuit with experiment, improving and fine-tuning it - is the second stage of the Blue Brain project, expected to be complete by the end of 2007.

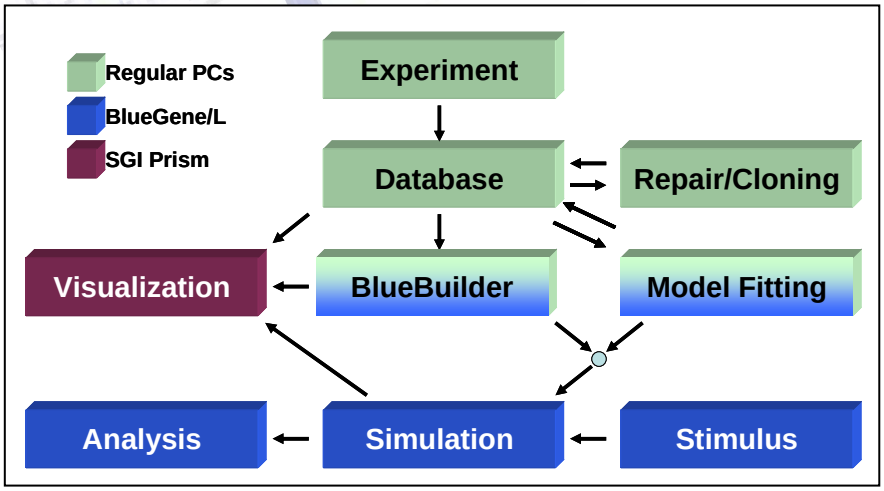

Figure 1: BBP Toolchain

Conclusions

Phase I marks the completion of a proof-of-principle simulation-based research process that has resulted in a cellular-level model of the neocortical column. We have achieved biological fidelity such that the model itself now serves as a primary tool for evaluating the consistency and relevance of neurobiological

data, while providing guidance for new experimental efforts. These new data will serve to further refine the neocortical column model. The assembled process allows neuroscientists to investigate scientific questions by integrating the available experimental data and evaluating hypotheses of network dynamics and neural function.

The completion of phase I provides the basis now for increasing the resolution of the models down to the molecular level and expanding the size of the models towards the whole brains of mammals.
In the future, information from the molecular and genetic level will be added to the algorithms that generate the individual neurons and their connections, and thus this level of detail will be reflected in the circuit's construction. The simulations can be used to explore what happens when this molecular or genetic information is altered -- situations such as a genetic variation in particular neurotransmitters, or what happens when the molecular environment is altered via drugs.

The project will continue to expand and will necessarily involve additional scientists and research groups from around the world.

UPM PARTICIPATION IN PHASE II

Although UPM has been involved in BBP since the early stages of the project, mainly in the Visualization tasks, the participation will be more representative in this new phase. The present proposal includes the open research lines in which UPM will take part.

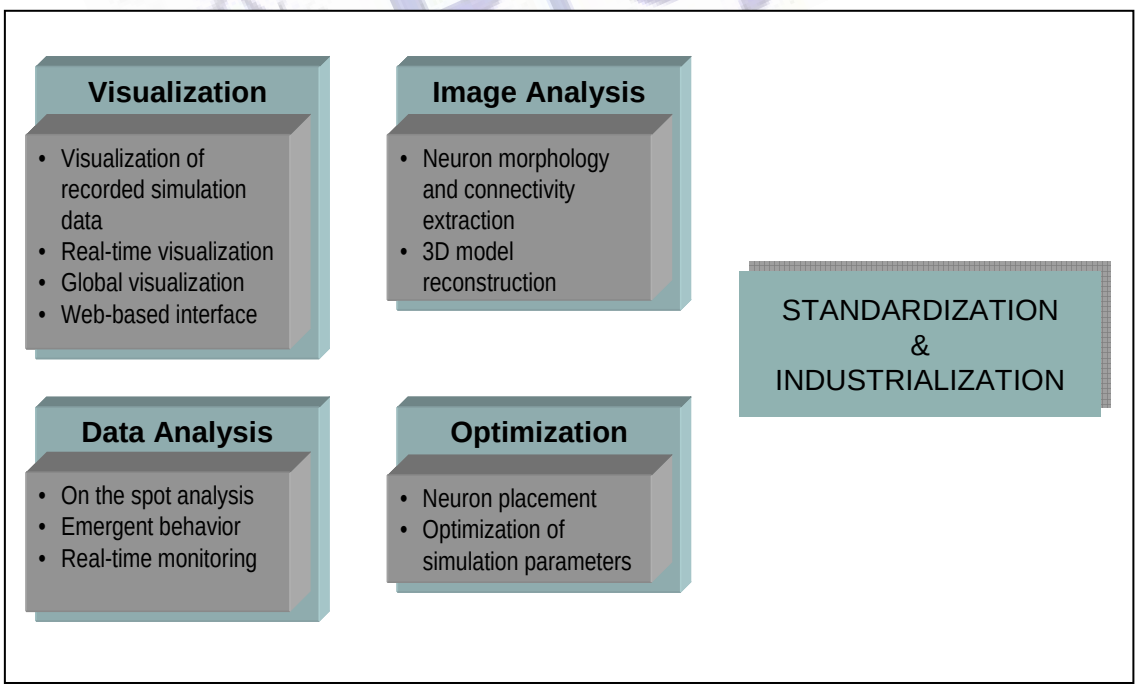

Figure 2: BB-UPM group working areas

The *Blue Brain - Universidad Politécnica de Madrid* group (BB-UPM group) will develop its project at the core of the BBP. The proposal will focus in the development and refinement of the Blue Brain informatics mainly in four different areas. Some of these areas are quite related between them.

Visualization

In the *Visualization* area the core developments will be focused in interactive visualization of simulation results. The main purposes of the software developed will be two-fold. On one hand, provide a tool for researchers to visualize interactively the large data sets generated in the BBP and on the other hand, support simulation run-time inspection. The first case deals with a post mortem scenario, where the main concern is high interactivity in the exploration of a large dataset. In both cases, remote visualization is a basic topic to be addressed.

Simulation analysis has will also be considered in this first area, allowing these processes to be interposed in the data path from simulation to visualization. In middle term, steering needs to be integrated in the whole data flow, to allow better stimulus-response analysis. The main issues involving the visualization infrastructure will be considered.

The developed software will benefit from IBM's DVC infraestructure as well as custom developments as needed, and including other software (overall open source licensed) from which we could benefit from.

Image Analysis

A second area will focus in *Image analysis for data capture*. We will develop new methods for computer-assisted 3D reconstruction of neurons from image stacks, actually a very time-demanding job. This task will help in the neuron morphology extraction as well as the connectivity extraction.

A second aspect of the research line on image analysis will come from the use of in-vivo activity areas based of brain response to different environmental stimuli. These studies have already been performed based on functional RMI imagery using different subjects. This approach will open the possibility to provide large-scale distant interaction areas to be considered together with the simulated scenarios.

Optimization

The *Optimization* area is directed towards the development and tuning of new parallel heuristic optimization techniques valid for BBP. The heuristic optimization techniques based in populations have already proved their usefulness in real problems. The possibilities provided by these techniques allow the simulation process to optimize different functional aspects. In the current approach, the optimization of some parameters can only be performed either by analytical optimization (if the function could be treated by some mathematical procedures) or by checking some possible values selecting the best of them (if the function complexity cannot be managed by analytical procedures). Heuristic optimization can achieve extensive optimization procedures in complex and large search spaces.

Within this context the new techniques will be used: (i) for optimizing neuron placement (i.e. maximization of structural unfiltered touches). This work will be done inside the BBP *BlueBuilder* module. (ii) For

optimizing calibration parameters. Having into account the expected outputs given specific inputs, it is possible to optimize the calibration parameters of the model.

Data Analysis

The last area will be related with the *Analysis of simulation results*. The objective is to develop a data analysis pipeline to extract high-level patterns of the simulation data. This pipeline is a complete procedure, which include some support tools to be developed within the project. These tools will provide (i) efficient data storage of the continuous data flow generated by the simulation (simulation data is in the scale of GB per simulated second), (ii) efficient data retrieval mechanism, and (iii) flexibility for the inclusion of new simulation records to perform both storage and querying.

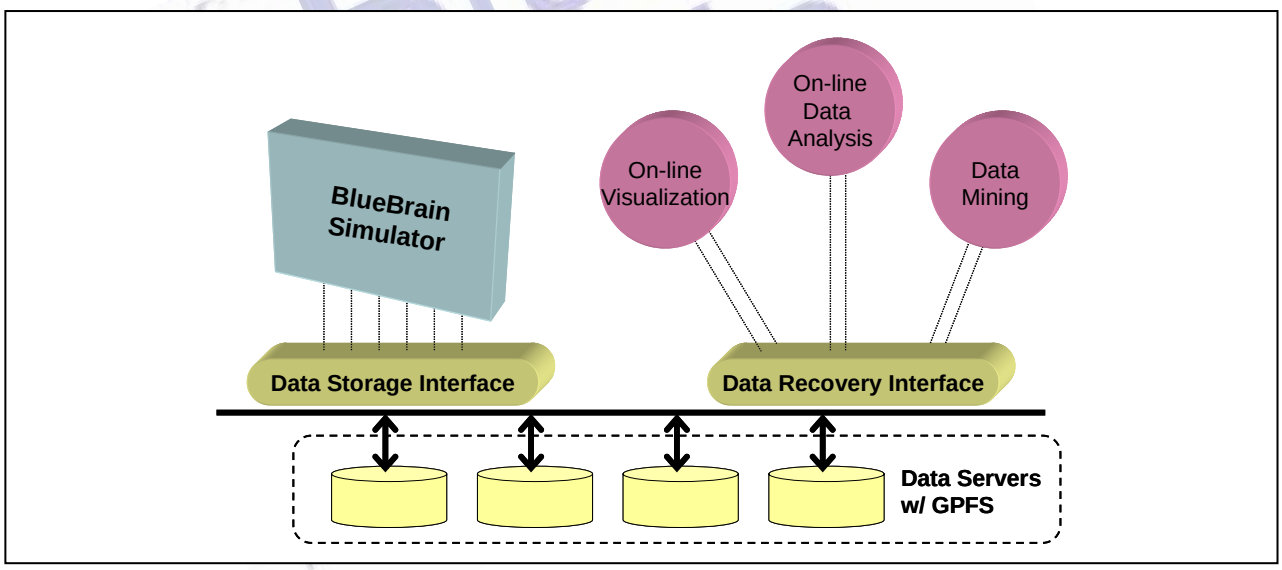

Figure 2: Data analysis pipeline

The next steps in the analysis pipeline concerns different intelligent data analysis techniques. These techniques will include (i) aggregated statistical information, (ii) interface with the data visualization software, and (iii) high-level pattern extraction techniques using data mining. This last group of techniques will provide the most interesting explanation of the emergent behaviour arisen from the simulation of these complex systems. Data mining can be one of the most powerful tools available for the BBP, allowing the extraction of hidden predictive and descriptive information, which can be of great interest for researches. Previously unknown information can be extracted using these on the spot analysis of simulation data.

Standardization

Finally, all the areas will require a process of standardization, which will be supported also by the BB-UPM group.

BENEFITS OF THE BLUEBRAIN PROJECT

Some of the potential benefits derived from Blue Brain project are [1].

Gathering and Testing 100 Years of Data

The most immediate benefit is to provide a working model into which the past 100 years knowledge about the microstructure and workings of the neocortical column can be gathered and tested. The Blue Column will therefore also produce a virtual library to explore in 3D the microarchitecture of the neocortex and access all key research relating to its structure and function.

Cracking the Neural Code

The Neural Code refers to how the brain builds objects using electrical patterns. In the same way that the neuron is the elementary cell for computing in the brain, the NCC is the elementary network for computing in the neocortex. Creating an accurate replica of the NCC which faithfully reproduces the emergent electrical dynamics of the real microcircuit, is an absolute requirement to revealing how the neocortex processes, stores and retrieves information.

Understanding Neocortical Information Processing

The power of an accurate simulation lies in the predictions that can be generated about the neocortex. Indeed, iterations between simulations and experiments are essential to build an accurate copy of the NCC. These iterations are therefore expected to reveal the function of individual elements (neurons, synapses, ion channels, receptors), pathways (mono-synaptic, disynaptic, multisynaptic loops) and physiological processes (functional properties, learning, reward, goal-oriented behaviour).

A Novel Tool for Drug Discovery for Brain Disorders

Understanding the functions of different elements and pathways of the NCC will provide a concrete foundation to explore the cellular and synaptic bases of a wide spectrum of neurological and psychiatric diseases. The impact of receptor, ion channel, cellular and synaptic deficits could be tested in simulations and the optimal experimental tests can be determined.

A Global Facility

A software replica of a NCC will allow researchers to explore hypotheses of brain function and dysfunction accelerating research. Simulation runs could determine which parameters should be used and measured in the experiments. An advanced 2D, 3D and 3D immersive visualization system will allow "imaging" of many aspects of neural dynamics during processing, storage and retrieval of information. Such imaging experiments may be impossible in reality or may be prohibitively expensive to perform.

A Foundation for Whole Brain Simulations

With current and envisageable future computer technology it seems unlikely that a mammalian brain can be simulated with full cellular and synaptic complexity (above the molecular level). An accurate replica

of an NCC is therefore required in order to generate reduced models that retain critical functions and computational capabilities, which can be duplicated and interconnected to form neocortical brain regions. Knowledge of the NCC architecture can be transferred to facilitate reconstruction of subcortical brain regions.

A Foundation for Molecular Modelling of Brain Function

An accurate cellular replica of the neocortical column will provide the first and essential step to a gradual increase in model complexity moving towards a molecular level description of the neocortex with biochemical pathways being simulated. A molecular level model of the NCC will provide the substrate for interfacing gene expression with the network structure and function. The NCC lies at the interface between the genes and complex cognitive functions. Establishing this link will allow predictions of the cognitive consequences of genetic disorders and allow reverse engineering of cognitive deficits to determine the genetic and molecular causes. This level of simulation will become a reality with the most advanced phase of Blue Gene development.

AVAILABLE INFORMATION

BlueBrain – Universidad Politécnica de Madrid (BB-UPM)
Campus de Montegancedo S/N,
Boadilla del Monte 28660 Madrid
Spain
Phone: +34 91 336 _____
Email: bb-upm@fi.upm.es
http://_____.fi.upm.es

REFERENCES

1. BlueBrain Project webpage: http://bluebrain.epfl.ch

2. Markram, H., 2006. *The blue brain project*. Nat Rev Neurosci. 7, 153-160.

3. Carey, J. (ed.), 2006. *Brain facts*. Society for Neuroscience 5th edition.