

Аналого-цифровые системы и высокопроизводительные решения в задачах исследования и моделирования человеческого мозга

Ананьева Н.И.¹, Богданов А.В.², Гушчанский Д.Е.², Дегтярев А.Б.², Залуцкая Н.М.¹, Лысов К.А.²,
Незнанов Н.Г.¹, Якушкин О.О.²

¹ ФГБУ «Санкт-Петербургский научно-исследовательский психоневрологический институт
им. В.М. Бехтерева» МЗ РФ,

² Санкт-Петербургский государственный университет

Резюме. Работа посвящена рассмотрению актуальных задач нейроинформатики в разрезе использования существующих аналого-цифровых систем и высокопроизводительных решений. Показаны и проанализированы существующие способы реализации и нужды различных экспериментов.

Отдельное внимание уделено задаче моделирования человеческого мозга на основе комбинирования цифровой и аналоговой подсистем. Рассмотрены существующие подходы моделирования взаимодействий нейронов мозга и используемые в них алгоритмы. В рамках рассмотрения способов применения облачных решений в задачах нейроинформатики выделены области особого интереса, такие как работа с внешними хранилищами, распределенная обработка информации и представление результатов. По каждой из них рассмотрены существующие опции для создания распределенных облачных вычислительных систем, решающих важные задачи, с которыми сталкиваются исследователи в ходе своей работы.

Ключевые слова: нейроинформатика, вычислительная нейробиология, высокопроизводительные вычисления, моделирование человеческого мозга

Analog and digital systems and high-performance solutions in problems of brain research and modeling

Ananyeva N.I.¹, Bogdanov A.V.², Gushchanskiy D.E.², Degtyarev A.B.², Zalutskaya N.M.¹, Lysov K.A.²,
Neznanov N.G.¹, Iakushkin O.O.²

¹ St.Petersburg V.M. Bekhterev Psychoneurological Research Institute,

² St. Petersburg University

Summary. This paper is devoted to some of the infrastructural problems of neuroinformatics in the context of existing analog-to-digital systems and high-performance solutions. We surveyed results and analyzed some of the existing methods required in various experiments.

Special attention was devoted to the problem of the human brain simulation, based on a combination of digital and analog subsystems. Paper shows various approaches to modeling of interactions between brain neurons and explores different algorithms.

We consider different applications of cloud solutions to neuroinformatics problems. Paper highlights areas of particular interest such as working with external storage, distributed data processing and visualization of acquired results. For each of them we survey existing options for creating cloud-based distributed computing solutions that can solve important challenges faced by researchers in their work.

Key words: neuroinformatics, computational neuroscience high performance computing, brain modelling.

Информационные технологии служат катализатором прогресса в любой научной дисциплине. Растущие вычислительные мощности, появление новых подходов, способных решать все более трудно формализуемые задачи, доступ к все большему количеству информации выводят исследования на новый уровень и служат отправной точкой для организации междисциплинарных исследований разного рода.

В складывающейся ситуации роль информационных технологий в исследованиях мозга особенно интересна тем, что данное сотрудничество нельзя назвать односторонним. Как и высокопроизводительные вычисления позволяют исследователям мозга обрабатывать и визуализировать огромные массивы данных о мозге, так и произ-

водятся попытки переложить результаты исследований устройства и работы мозга на вычислительную платформу, тем самым существенно расширив круг решаемых информационными технологиями задач.

1. Моделирование человеческого мозга на основе комбинирования цифровой и аналоговой подсистем

Последнее десятилетие характеризуется резким ростом интереса к нейроморфным вычислениям, в частности, к моделированию человеческого мозга и особенностям его работы при принятии решений. Об этом свидетельствуют запуски масштабных исследовательских программ — DARPA

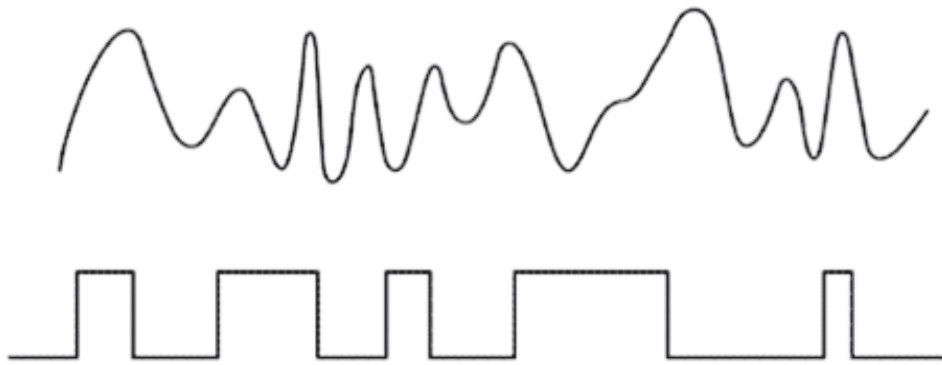


Рис. 1. Слева — аналоговый сигнал, справа — цифровой.

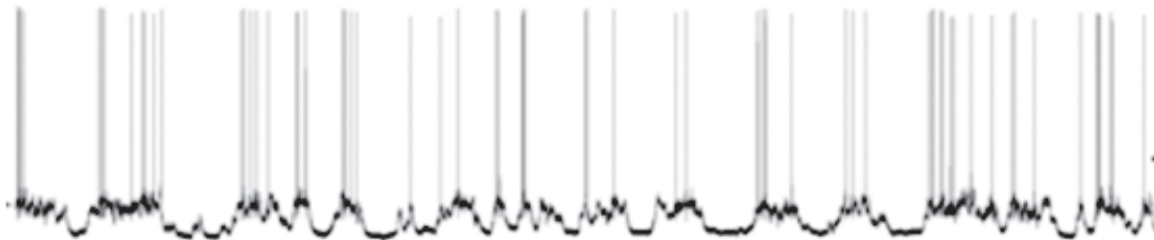


Рис. 2. Нейронный сигнал.

SyNAPSE (США) [1] и Human Brain Project (ЕС) [2], целью которых является построение микропроцессорной системы, которая повторяет мозг человека по функционалу, размеру и потребности энергии.

Существующие модели мозга даже на мощных суперкомпьютерах требуют значительного времени вычислений и пока не могут решать задачи реального времени. Поскольку человеческий мозг состоит из двух частей с разными функциями и разными принципами обработки данных, то весьма перспективным является подход, объединяющий в единый комплекс цифровую и аналоговую подсистемы. Для его применения необходимо уделить исключительное внимание архитектуре аналоговой системы и ее совместимости с цифровыми решениями.

1.1 Кодирование в нейронах мозга

Один из крупнейших споров нейробиологии ведется вокруг того, как нейроны кодируют информацию. Неясно, посылается ли информация в цифровой или аналоговой форме (рис. 1), либо мозг пользуется обоими средствами одновременно (рис. 2).

Цифровые сигналы, передаваемые обычными компьютерами, нисколько не похожи на аналоговые сигналы, применяемые в старых телевизорах и радиоприемниках. Отличить их друг от друга просто, чего нельзя сказать того же о нейронных сигналах — там разделить цифровые и аналоговые сигналы довольно сложно. Поэтому целесообразно

было бы найти способ определения того, является ли нейронный сигнал аналоговым или цифровым.

Нейробиологам давно известно, что нейроны передают сигналы в форме электрических импульсов, которые называются биоэлектрическими потенциалами или «спайками». Несколькими взятыми вместе спайков называется последовательностью спайков. Точный способ кодирования информации в спайках неизвестен, однако ученые открыли как минимум два протокола кодирования. В 1990 году нейробиологи обнаружили, что напряжение мышцы зависит от количества «спайков» в определенный период времени, от скорости их прибывания. Этот вид сигнала имеет лишь два состояния — включено или выключено — так что он определенно является цифровым. Однако другие нейробиологи утверждают, что информация может быть закодирована и по-другому — посредством разницы во времени между отдельными спайками при их прибытии. Это аналоговое кодирование.

Сложность заключается в разграничение этих двух сигналов, поскольку они оба зависят от характеристики спайков, которые путешествуют по нейрону. Этот вопрос вызывает частые споры среди нейробиологов, поскольку отсутствует согласия относительно того, когда сигнал является цифровым, а когда аналоговым.

Не так давно японские физики Ясухио Мотидзуки (Yasuhiro Mochizuki) и Сигеру Синомото (Shigeru Shinomoto) из Университета Киото разработали способ автоматического определения вида

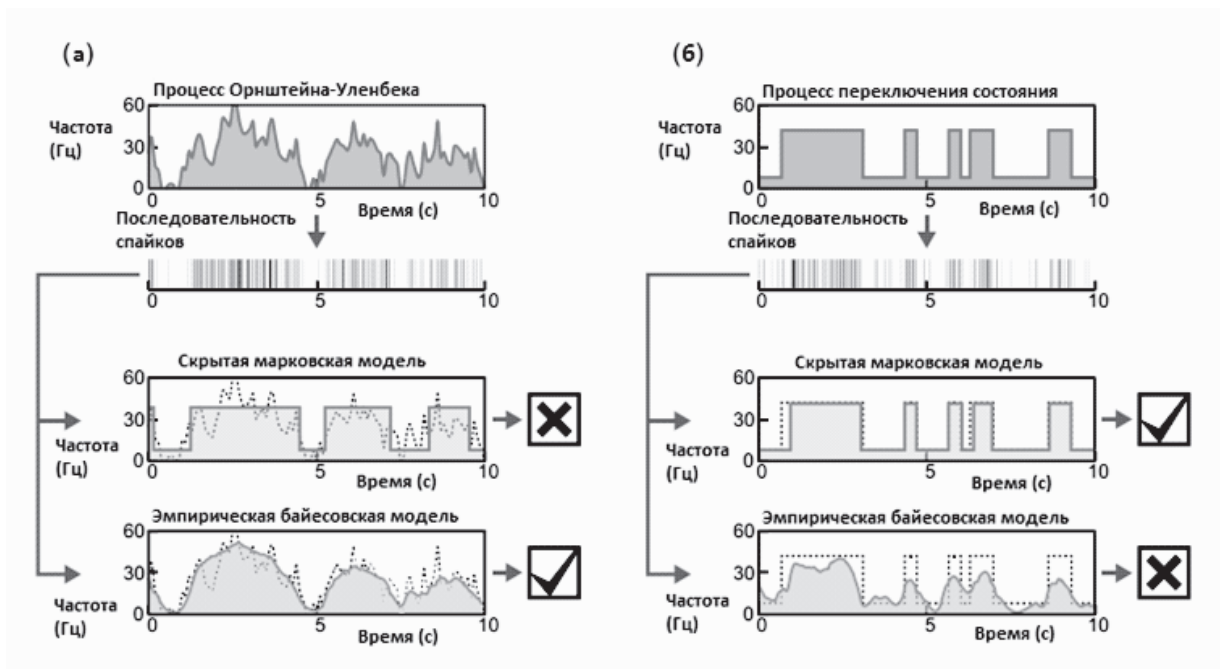


Рис. 3. (а) Последовательность спайков, сгенерированная с помощью процесса Орнштейна-Уленбека (Ornstein-Uhlenbeck Process) (синий). Эмпирическая байесовская модель (зеленый) лучше аппроксимирует сигнал, чем скрытая марковская модель (оранжевый). (б) Последовательность спайков, полученная с помощью процесса переключения состояния (Switching State process) (красный). Скрытая марковская модель лучше аппроксимирует сигнал, чем эмпирическая байесовская модель.

кодирования [3]. Способ основан на идее о том, что некоторые статистические модели лучше выражают цифровой код, чем аналоговый код, и наоборот.

Метод довольно прямолинеен. Ученые анализируют сигнал нейрона и затем стараются повторить его сначала с помощью эмпирической байесовской модели, а затем — с помощью скрытой марковской модели. Далее на основе модели, которая лучше отражает характеристики первоначального сигнала, они определяют, является ли сигнал аналоговым или цифровым. Получается, что если эмпирическая байесовская модель лучше отражает сигнал, тогда сигнал, вероятно, аналоговый, если же скрытая марковская модель подходит лучше, тогда сигнал, скорее всего, цифровой (рис. 3) [3].

Данный подход был проверен на сигналах, которые возникали в разных частях мозга длиннохвостых макаков, и подтвердил, что разные части мозга используют разные формы кодирования. Это дает повод проверить на практике обратную ситуацию, когда будет построена система, комбинирующая аналоговые и цифровые сигналы для генерации нейронных спайков.

1.2 Архитектура аналоговой подсистемы

В последние годы в рамках DARPA SyNAPSE [1] был разработан ряд нейроморфных (грубо повторяющих структуру нейронов и синапсов в мозгу человека) архитектур, реализующих концепцию резистивных процессорных устройств (Resistive Processing Unit, RPU) [4].

RPU — вычислительный элемент, аналоговый по своей природе, небольшой по размерам и способный восстанавливать свою историю, чтобы обучаться. Он получает множество аналоговых данных, в форме напряжений, и на основе прошлого опыта использует взвешенную функцию из них, чтобы решить, какой результат передавать на следующий слой вычислительных элементов (рис. 4). Синапсы имеют озадачивающее и пока непонятное положение в мозгу человека, но чипы из RPU организованы в двумерные массивы.

Из-за того, что RPU специализированы и не требуют преобразования аналоговой информации в цифровую или доступа к какой-либо памяти, кроме своей собственной, они могут быть быстры и поглощать мало энергии. Поэтому, теоретически, сложная нейронная сеть может быть напрямую смоделирована путем выделения одного RPU к одному программному нейрону. К сожалению, RPU неточен из-за своей аналоговой природы и обилия шума в схемах, поэтому алгоритм должен иметь устойчивость к «врожденным» неточностям в RPU.

2. Облачные решения в задачах нейроинформатики

Задачи нейроинформатики сконцентрированы на создании, хранении, обработке и визуализации результатов исследования. Все эти стадии затрагивают работу с большими объемами данных и требуют разработки уникального программного обеспечения, требующего специальных алго-

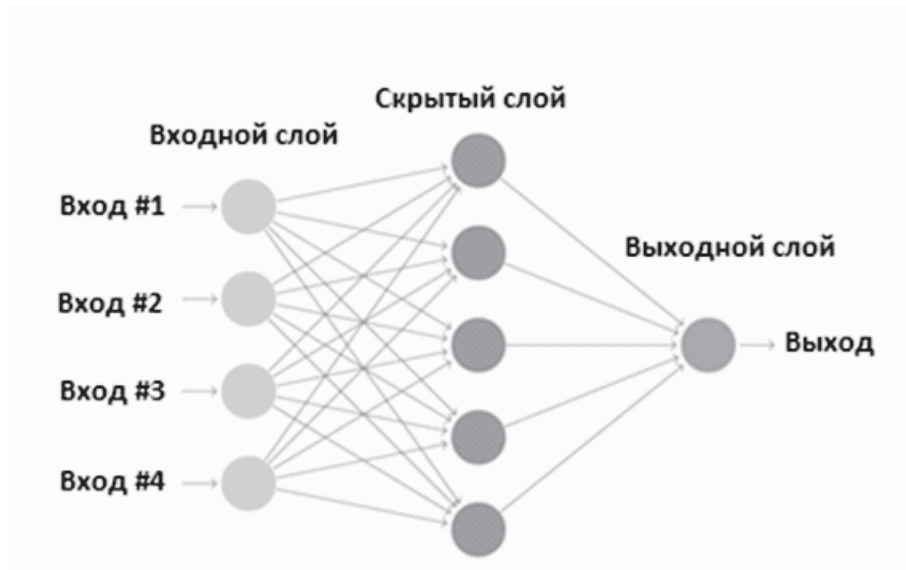


Рис. 4. Условная схема сети из RPU.

ритмов для эффективной работы. Сейчас различные компоненты облачной инфраструктуры уже используются в нейроинформатике, например, в таких проектах как JuBrain [5] и A-Brain [6]. Созданием облачных платформ для обеспечения работы в этой области активно занимались с 2008го года [7], однако системы, которая бы покрывала полный спектр задач, до сих пор не существует, и зачастую для решения отдельных задач требуется специализированные проекты [8].

2.1 Хранение и предобработка информации

Корректная запись информации нуждается в возможности черновой предобработки данных. Источниками таких данных могут быть аппараты MRI, EEG и другие, непосредственно взаимодействующие с объектом исследования. Анализ потоков данных, получаемых с этих аппаратов, требует мощной вычислительной системы. Так, например, одной из задач проекта fMRI является получение трехмерной модели активности мозга в реальном времени. В нем используются сопроцессорные технологии Xeon Phi [9]. Задействование кластерных систем напрямую ограничивает возможности перераспределения вычислительных ресурсов [10]. В таких системах ресурсы выделяются в рамках узкого спектра программных платформ, что ограничивает гибкость используемого программного обеспечения [11]. Решением таких проблем может стать частное облако, развернутое в среде исследовательского центра и позволяющее использовать виртуальные машины или Docker контейнеры [12, 13].

Рассмотрим технологии, создающие сверхбольшие объемы данных. Так, например, Гарвардский проект «CONNECTOME» требует анализа 2PB данных для каждого кубического миллиметра мозга [14]. Такие проекты находятся на ранней

стадии своего развития, однако приносят плодотворные результаты [15, 16] и показывают значительные успехи в области увеличения скорости обработки информации с двух столетий для одного миллиметра в 2010м году к 5.6 года в 2013-м [17]. Такие объемы данных требуют длительного хранения. Публичные облачные системы предоставляют возможности «холодного» хранилища [18], позволяющие размещать данные на медленных жестких дисках большого объема или магнитной пленке. Это позволяет решить задачу расширения инфраструктуры частного кластера, превращая его в гибридное облако, способное разместить необходимый объем данных. Аналогично удаленному хранилищу может осуществляться расширение «по требованию» вычислительных ресурсов «домашнего» облака.

2.1 Обработка данных

Помимо средств ручной агрегации данных существует ряд проектов, направленных на создание общей базы знаний. Проект института Аллена Brain Atlas агрегирует открытые данные, предоставляя программный интерфейс к изображениям, 3D реконструкциям, графам и структурам по направлениям изучения мозга: мышей и различных типов клеток, связей мозга. Эта база данных размещена в США [19]. Для проведения исследований с использованием этой базы в любом другом регионе, необходимо передавать данные на большие расстояния. Организация межрегиональной сети может быть не оптимальна, в силу своей динамической организации и постоянных изменений. Публичные облачные платформы позволяют строить защищенные каналы связи, которые имеют гарантированную пропускную способность для передачи данных между регионами.

Разработка средств обработки данных неотъемлемо связана с развитием высокопроизводительных вычислений. Так, в работе [20] представлен ряд систем моделирования нейронных сетей, использующих вычисления на сопроцессорах NVidia, доступных в публичных облачных решениях. В [21] рассматривается связь работы мозга и его моделирования с использованием квантовых вычислений, которые, как утверждают авторы [22], станут, в первую очередь, доступны в форме платформы-виде-сервиса от облачных провайдеров.

Много внимания уделяется системам, позволяющим работать в рамках модели “notebooks”. Такие системы направлены на объединение веб-сервиса, содержащего IDE, системы визуализации в формате HTML5 и возможности постановки задач на удаленные кластеры. В рамках использования таких систем для решения задач нейроинформатики отмечают решения, использующие платформу Spark [23]: Bolt для работы с локальными и распределенными массивами данных; Thunder для анализа изображений и временных рядов. Облачные поставщики часто выделяют в ряду своих разработок отдельные системы, позволяющие быстро развернуть Spark кластер.

2.3 Визуализация

Визуализация материалов исследования может служить ключом к научному открытию.

Проект The Virtual Brain [24] позволяет отображать различные аспекты, связанные с моделированием мозга в рамках единого фреймворка в виде веб-сервиса. Много-арендная модель рассматривает поддержку изолированной работы нескольких организаций в рамках единого нижележащего оборудования облачных центров обработки данных. Такое решение дает возможность одному поставщику системы визуализации полностью разделить данные клиентов, обеспечивая повышенную безопасность работы системы.

Облачные системы позволяют осуществлять максимальную экономию задействованных ресурсов в каждый конкретный момент времени. Публичные облачные решения расширяют имеющиеся мощности организации, задействовав технологии, недоступные в рамках локальных центров обработки данных. Это ускоряет разработку прототипов программного обеспечения и расширяет возможности существующей ИТ-инфраструктуры. Достоинством облачных решений является то, что от использования облачного оборудования можно отказаться в любой момент, изменить его параметры или расширить используемый объем за небольшой объем времени. Именно эти особенности облачных технологий делают их актуальными для решения задач нейроинформатики.

Литература

1. *Broad Agency Announcement. Systems of Neuromorphic Adaptive Plastic Scalable Electronics. DARPA-BAA 08-28. 9 April 2008.* — <https://www.fbo.gov/download/0b6/0b62b2149395d4bd8a28dff1b9046944/BAA08-28.doc>
2. *The Human Brain Project. A Report to the European Commission.* — https://ec.europa.eu/research/participants/portal/doc/call/h2020/fetflag-1-2014/1595110-6pilots-hbp-publicreport_en.pdf
3. Yasuhiro Mochizuki, Shigeru Shinomoto. *Analog and digital codes in the brain.* Department of Physics, Kyoto University, Kyoto 606-8502, Japan. November 19, 2013. — <http://arxiv.org/pdf/1311.4035v1.pdf>
4. Tayfun Gokmen, Yurii Vlasov. *Acceleration of Deep Neural Network Training with Resistive Cross-Point Devices.* IBM T. J. Watson Research Center. — <https://arxiv.org/ftp/arxiv/papers/1603/1603.07341.pdf>
5. Mohlberg, H., Eickhoff, S.B., Schleicher, A., Zilles, K. and Amunts, K. *A new processing pipeline and release of cytoarchitectonic probabilistic maps—JuBrain.* — 2012.
6. Antoniu, G., Costan, A., Mota, B.D., Thirion, B. and Tudoran, R. *A-brain: using the cloud to understand the impact of genetic variability on the brain.* — *ERCIM News*, 89. — 2012. — P. 21-22.
7. Watson, P., Lord, P., Gibson, F., Periorellis, P. and Pitsilis, G. *Cloud Computing for e-Science with CARMEN.* In *2nd Iberian Grid Infrastructure Conference Proceedings* — 2008. — May. — P. 3-14.
8. D’Haese, P.F., Konrad, P.E., Pallavaram, S., Li, R., Prasad, P., Rodriguez, W. and Dawant, B.M.. *CranialCloud: a cloud-based architecture to support trans-institutional collaborative efforts in neurodegenerative disorders.* — *International journal of computer assisted radiology and surgery.* — 2015. — Vol.10. — P. 815-823.
9. Wang, Yida, Michael J. Anderson, Jonathan D. Cohen, Alexander Heinecke, Kai Li, Nadeethur Satish, Narayanan Sundaram, Nicholas B. Turk-Browne, and Theodore L. Willke. *«Full correlation matrix analysis of fMRI data on Intel® Xeon Phi™ coprocessors.»* In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis.* — ACM. — 2015. — p. 23.
10. Богданов А.В., Дегтярев А.Б. *Суперкомпьютинг без суперкомпьютеров: что мы можем, а что не можем? Вычислительные технологии в естественных науках. Методы суперкомпьютерного моделирования. Сборник трудов. Сер. «Механика, управление и информатика» под редакцией П.Р. Назирова, Л.Н. Шура.* — Институт космических исследований Российской академии наук. — Москва. — 2015. — С. 61-77.
11. Bogdanov, A., Degtyarev, A. and Korkhov, V. *New Approach to the Simulation of Complex Systems.*

- In EPJ Web of Conferences. — 2016. — Vol. 108. — P. 01002.
12. Jinzhou, Yang, He Jin, Zhang Kai, and Wang Zhi-jun. «Discussion on private cloud PaaS construction of large scale enterprise.» In 2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). — 2016. — P. 273-278.
 13. V. Korkhov, I. Gankevich, A. Degtyarev, A. Bogdanov, V. Gaiduchok, N. Ahmed, A. Cubahiro. “Experience in Building Virtual Private Supercomputer”, Proceedings of 13. International Conference on Computer Science and Information Technologies (CSIT). — 2015. — P. 220-223. — ISBN 978-5-8080-0797-0
 14. Swanson, L.W. and Lichtman, J.W. From Cajal to Connectome and Beyond. — Annual Review of Neuroscience. — 2016. — Vol. 39.
 15. Tomassy, G.S., Berger, D.R., Chen, H.H., Kasthuri, N., Hayworth, K.J., Vercelli, A., Seung, H.S., Lichtman, J.W. and Arlotta, P. Distinct profiles of myelin distribution along single axons of pyramidal neurons in the neocortex. — Science. — 2014. — Vol. 344. — P. 319-324.
 16. Lichtman, J.W. and Denk, W. The big and the small: challenges of imaging the brain's circuits. — Science. — 2011. — Vol. 334. — P. 618-623.
 17. Lichtman, J.W., Pfister, H. and Shavit, N. The big data challenges of connectomics. — Nature neuroscience. — 2014. — Vol.17. — P. 1448-1454.
 18. Han, Y. Cloud storage for digital preservation: optimal uses of Amazon S3 and Glacier. — Library Hi Tech. — 2015. — Vol. 33. — P. 261-271.
 19. Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K. and Arnold, J.M. Transcriptional landscape of the prenatal human brain. — Nature. — 2014. — Vol.508. — P. 199-206.
 20. Prieto, A., Prieto, B., Ortigosa, E.M., Ros, E., Pelayo, F., Ortega, J. and Rojas, I. Neural networks: An overview of early research, current frameworks and new challenges. — Neurocomputing. — 2016.
 21. Neven, H., Denchev, V.S., Rose, G. and Macready, W.G. QBoost: Large Scale Classifier Training with Adiabatic Quantum Optimization. — In ACML. — 2012. — P. 333-348.
 22. Singh, H. and Sachdev, A., 2014, February. The quantum way of cloud computing. In Optimization, Reliability, and Information Technology (ICROIT). — International Conference on. — 2014. — P. 397-400. Ieee.
 23. Cunningham, J.P. Analyzing neural data at huge scale. — Nature methods. — 2014. — Vol.11. — P. 911-912.
 24. Leon, P.S., Knock, S.A., Woodman, M.M., Domide, L., Mersmann, J., McIntosh, A.R. and Jirsa, V. The Virtual Brain: a simulator of primate brain network dynamics. Information-based methods for neuroimaging: analyzing structure, function and dynamics. — 2015. — P. 10.

Сведения об авторах

Ананьева Наталия Исаевна — д. м. н., профессор, руководитель отделения клинко-диагностических исследований ФГБУ «Санкт Петербургский научно-исследовательский психоневрологический институт им. В.М. Бехтерева». E-mail: ananieva_n@mail.ru.

Богданов Александр Владимирович — д. ф.-м. н., профессор кафедры компьютерного моделирования и многопроцессорных систем СПбГУ. E-mail: bogdanov@csa.ru

Гущанский Дмитрий Евгеньевич — ассистент кафедры компьютерного моделирования и многопроцессорных систем СПбГУ. E-mail: dmitriy.guschanskiy@spbu.ru

Дегтярев Александр Борисович — д. т. н., профессор кафедры компьютерного моделирования и многопроцессорных систем СПбГУ. E-mail: deg@csa.ru

Залуцкая Наталья Михайловна — к. м. н., доцент, ведущий научный сотрудник отделения гериатрической психиатрии ФГБУ «Санкт Петербургский научно-исследовательский психоневрологический институт им. В.М. Бехтерева». E-mail: nzalutskaya@yandex.ru

Лысов Кирилл Александрович — студент СПбГУ. E-mail: thereis9000@gmail.com

Незнанов Николай Григорьевич — д. м. н., профессор, директор ФГБУ «Санкт-Петербургский научно-исследовательский психоневрологический институт им. В.М. Бехтерева», научный руководитель отделения гериатрической психиатрии ФГБУ «Санкт-Петербургский научно-исследовательский психоневрологический институт им. В.М. Бехтерева». E-mail: spbinstb@bekhterev.ru

Якушкин Олег Олегович — ассистент кафедры компьютерного моделирования и многопроцессорных систем СПбГУ. E-mail: oleg.jakushkin@gmail.com