

**Нейросетевые технологии в перспективных суперЭВМ. Концепция развития высокопроизводительных вычислений на базе супернейрокомпьютеров (2012-2020г.г.)**

*Галушкин А.И.*

**Аннотация:** Представлена история развития масштабируемых суперЭВМ с двухслойной архитектурой и их оценка с точки зрения эффективности моделирования сложных нейронных сетей. Отмечается ведущая роль в настоящее время суперЭВМ на базе графических процессоров в проблеме моделирования нейронных сетей. Представлена концепция развития супернейрокомпьютеров – суперЭВМ на базе нейросетевых технологий.

**Ключевые слова:** нейросетевые технологии, супернейрокомпьютеры, двухслойные архитектуры, масштабируемость.

**Neural network technology in perspective supercomputers (development concept of supercomputers on the base of neural network technology in 2012-2020)**

*Galushkin A.I.*

**Annotation.** The history of development of scaled super computers with two-layer architecture and their assessment from the point of view of efficiency of modeling of difficult neural networks is presented. The leading role by the supercomputer on the basis of graphic processors in a problem of modeling of neural networks is noted now. The concept of development of superneurocomputers – the super computer on the basis of neural network technologies is presented.

**Keywords:** neural network technologies, superneurocomputers, two-layer architecture, scalability.

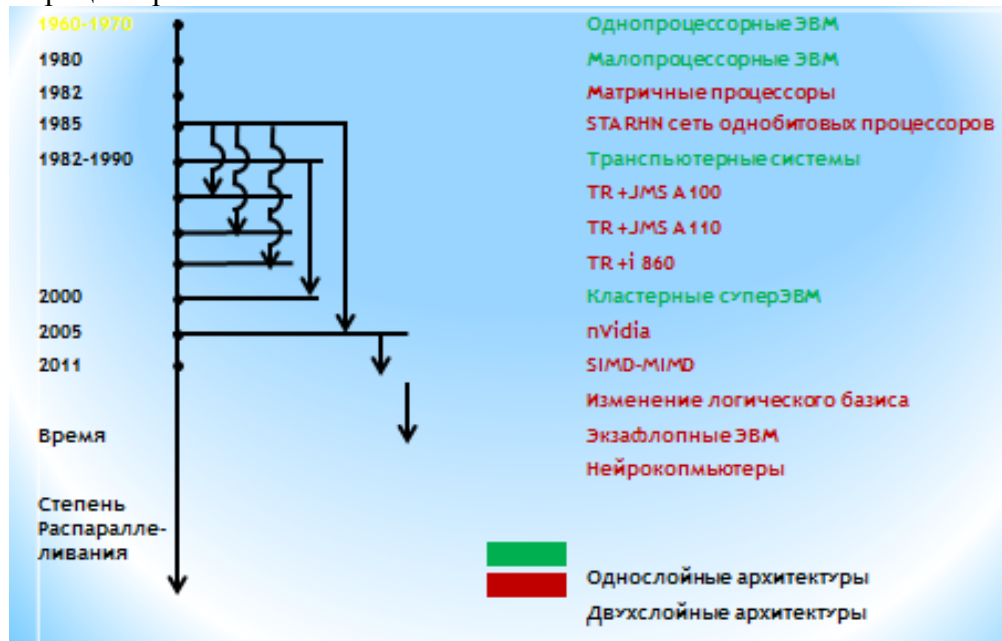
Данная работа посвящена краткому изложению развития масштабируемых суперЭВМ с двухслойной архитектурой. Эта история схематично представлена на рис. 1, где отмечено, что первыми образцами ЭВМ с двухслойной архитектурой были ЭВМ с матричными сопроцессорами и суперЭВМ типа STARAN.

В соответствие с профессиональными интересами автора большинство типов высокопроизводительных вычислительных систем исследовалось тестированием моделями сложных нейронных сетей. Именно этот тест был главным при оценке производительности ЭВМ той или иной архитектуры. Количественным критерием производительности являлись и являются в настоящее время следующие три показателя:

- количество эмулируемых нейронов;
- количество эмулируемых связей между нейронами;
- скорость переключения связей в секунду.

Транспьютерные системы [1] стали предвестником многих современных суперЭВМ, как кластерных, так и гибридных. Транспьютерные системы кластерного типа оказались неэффективными для моделирования сложных нейронных сетей, однако еще в конце 80-ых – начале 90-ых годов прошлого столетия эффективными для решения этой задачи были предвестники современных гибридных суперЭВМ, а именно двухслойные транспьютерные архитектуры (рис. 2) с периферийными [1]:

- сигнальными процессорами JMSA100;
- процессорами обработки изображений JMSA110;
- процессорами i860

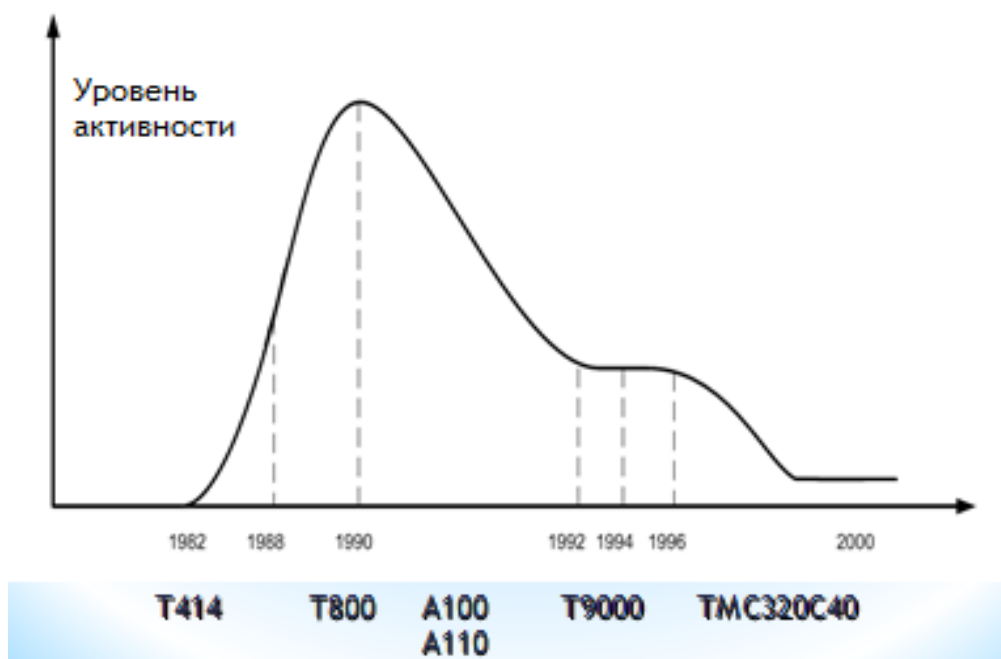


**Рисунок 1. Тенденции развития архитектуры суперЭВМ.**

Гибридные суперЭВМ на базе процессоров nVidia являются, по сути дела. Гармоничным развитием этой архитектуры на более совершенном технологическом уровне. Если отметить два основных свойства современных суперЭВМ:

- масштабируемость,
- двухслойность архитектур,

то, с нашей точки зрения, для перехода на экзафлопный уровень вычислений необходима реализация, по крайней мере, еще одного шага – изменения логического базиса алгоритмов решения задач и соответственно элементной базы.



**Рисунок 2. Активность разработок и применения транспьютерных систем**

В [2] отмечено следующее: «Белый дом заявил о нежелании втягиваться в гонку вооружений в вопросе создания все более быстрых компьютеров, и год назад в своем докладе предупредил, что акцент на скорость «может отвлечь ресурсы от фундаментальных исследований, направленных на развитие принципиально новых подходов, которые помогли бы нам, в конечном итоге, обогнать другие страны»».

Основными предпосылками принципиального изменения архитектуры вычислительных систем при переходе к системам экзафлопной производительности являются следующие:

- Необходимость резкого повышения надежности за счет отказа от фон-Неймановской архитектуры вычислительных систем из элементов булевской логики И, ИЛИ, НЕ, когда в системе из значительного числа этих элементов происходит катастрофический отказ при отказе любого элемента;

- Необходимость резкого снижения энергопотребления за счет отказа от существующего позиционного метода представления информации и перехода к другим, обеспечивающим снижение энергопотребления и как следствие, дополнительное повышение надежности;

- Необходимость резкого повышения однородности схемотехники элементов вычислительных систем, что должно привести к повышению эффективности при том резком увеличении интеграции элементов, которое будет иметь место при переходе к экзафлопной производительности.

Детальный анализ документа [3] показал необходимость разработки подобного документа, ориентированного на использование при разработке экзафлопных суперЭВМ нейросетевых технологий [4].

К разработке документа [4] «Экзафлопные нейросетевые технологии. Концепция развития технологий высокопроизводительных вычислений на базе супернейрокомпьютеров (2012-2020 гг.)» было привлечено более сотни ведущих российских ученых в области нейросетевых технологий, в основном работающих в российских университетах с некоторым, к сожалению незначительным, участием ученых из РАН.

Научно-техническими заделами по созданию отечественных супернейрокомпьютеров экзафлопной производительности являются заделы по следующим направлениям:

- Теория нейронных сетей, как методика синтеза структур из нейронов различного вида, алгоритмов адаптации весовых коэффициентов в этих структурах в процессе решения различных задач,

- Нейроматематика, как раздел вычислительной математики, связанный с решением в нейросетевом логическом базисе различных сложных формализуемых и неформализуемых задач,

- Нейроуправление, как раздел теории управления, связанный с применением нейрокомпьютеров в качестве систем идентификации сложных динамических систем и нейрокомпьютеров,

- Нейрокомпьютеры и нейрочипы.

В работе [5] представлен систематизированный материал аннотаций работ российских авторов в области развития и применения нейросетевых технологий за 1982-2010 года, а именно:

- диссертационные работы (кандидатские и докторские);

- монографии;

- научно-технические отчеты фонда ВНИИЦ.

Анализ этих материалов показывает значительный потенциал ученых в России в области развития и применения нейросетевых технологий.

Известны ведущие в мире основные проекты супернейрокомпьютеров экзафлопной производительности:

- Проект Irvine Sensors;
- Проект Ливерморской лаборатории;
- Проект Института продвинутых архитектур (Сандиа, Оак Ридж);
- Проект Интел, центр суперкомпьютерных вычислений Сан-Диего, DARPA;
- Проект eLiza - самоуправляемой автономной компьютерной системы (IBM);
- Проект SyNAPSE (Systems Neuromorphic Adaptive Plastic Scalable Electronics);
- Проект Neurogrid (Стэнфордский университет)
- Проект SpiNNaker (Spiking Neural Network Architecture) – группа английских университетов;
- BICA (Biologically Inspired Computer Architecture);
- Проект BlueBrain.

Основными прикладными задачами, стимулирующими развитие суперЭВМ, в том числе супернейрокомпьютеров, являются следующие:

- Биоинформатика;
- Криптография;
- Ядерные исследования;
- Газо-, аэро-, гидродинамика;
- Обработка космических изображений;
- Обработка больших массивов текстовой информации;
- Моделирование мозга.

В работе [6] был представлен предварительный рабочий материал к Концепция развития технологий высокопроизводительных вычислений на базе супернейрокомпьютеров. В [4] и [6] отмечены:

**1. Задачи, требующие разработки и применения нейросетевых технологий:**

- 1.1. Нейросетевые технологии в хемо- и биоинформатике,
- 1.2. Нейросетевые технологии в криптографии,
- 1.3. Нейросетевые технологии в ядерной энергетике,
- 1.4. Нейросетевые технологии в медицине ,
- 1.5. Нейросетевые технологии в системах добычи и обработки полезных ископаемых,
- 1.6. Нейросетевые технологии в управлении энергетическими системами,
- 1.7. Нейросетевые технологии в управлении транспортными системами,
- 1.8. Нейросетевые технологии в биометрии,
- 1.9. Нейросетевые технологии в обработке статических и видеоизображений ,
- 1.10. Нейросетевые технологии в обработке сигналов,
- 1.11. Нейросетевые технологии в обработке больших массивов текстовой информации,
- 1.12. Нейросетевые технологии в разработках интерфейсов мозг-компьютер,
- 1.13. Нейросетевые технологии в авиации,
- 1.14. Нейросетевые технологии в космической технике,
- 1.15. Нейросетевые технологии в роботостроении,
- 1.16. Нейросетевые технологии в моделировании сложных технических систем,
- 1.17. Нейросетевые технологии в распределённых вычислениях и моделировании распределённых систем,
- 1.18. Нейросетевые технологии в экологии,

1.19. Нейросетевые модели экономических, социальных, политических процессов;

**2. Направления фундаментальных исследований в области нейросетевых технологий:**

2.1. Теория нейронных сетей,

2.2. Нейросетевые технологии в решении сложных математических задач (нейроматематика),

2.3. Нейроуправление,

2.4. Разработка моделей разделов мозга для современных и перспективных суперЭВМ;

**3. Направления, связанные с созданием супернейрокомпьютеров:**

3.1. Нейрокомпьютеры,

3.2. Основные предпосылки предлагаемого направления развития работ по архитектуре перспективных эксафлопных суперЭВМ с применением нейросетевых технологий,

3.3. Нейрочипы:

3.3.1. Цифровые нейрочипы,

3.3.2. Аналоговые и аналогово-цифровые нейрочипы,

3.3.3. Супернейрокомпьютер фирмы Irvine Sensors,

3.3.4. Мемристоры,

3.3.5. Клеточные нейрочипы,

3.3.6. Нейрочипы с частотно-импульсной модуляцией сигналов,

3.3.7. Специализированные аналоговые и аналого-цифровые нейрочипы,

3.4. Перспективные технологии в реализации супернейрокомпьютеров:

3.4.1. Многослойные системы,

3.4.2. Оптические нейрокомпьютеры,

3.4.3. Одноэлектронные системы,

3.4.4. Молекулярные и ДНК-вычисления,

3.4.5. Квантовые нейрокомпьютеры,

3.5. Специализированные и периферийные системы:

3.5.1. Ассоциативная память,

3.5.2. Решатели систем линейных алгебраических уравнений,

3.5.3. Специализированный процессор обработки изображений,

3.6. Разработка системного программного обеспечения:

3.6.1. Операционная система,

3.6.2. Система ввода\вывода,

3.6.3. Система управления,

3.6.4. Внешняя среда,

3.6.5. Модель программирования,

3.6.6. Компиляторы,

3.6.7. Библиотеки,

3.6.8. Средства отладки,

3.7. Перспективные свойства супернейрокомпьютеров:

3.7.1. Ассоциативная память,

3.7.2. Эффективное распараллеливание алгоритмов решения задач,

3.7.3. Реконфигурируемость,

3.7.4. Масштабируемость,

3.7.5. Надежность,

3.7.6. Визуализация,

- 3.7.7 Облачные вычисления,
- 3.7.8 Тестирование программного обеспечения,
- 3.7.9 Методика оценки производительности нейрокомпьютеров,

### 3.8. Создание и развитие супернейрокомпьютерных центров.

Необходимо отметить объективные причины формирования представленной Концепции по развитию технологий высокопроизводительных вычислений на базе супернейрокомпьютеров (2012-2020 гг.):

1. В Концепции экзафлопных ЭВМ [3] отсутствует даже упоминание о нейросетевых технологиях.

2. В авторском коллективе Концепции экзафлопных ЭВМ [3] отсутствуют специалисты по нейросетевым технологиям.

3. Список цитируемой литературы в Концепции экзафлопных ЭВМ [3] говорит о слабости, если не сказать точнее о нулевой информированности ее разработчиков об активном движении зарубежных и отечественных ученых в направлении разработок сверхвысокопроизводительной вычислительной техники с применением нейросетевых технологий.

Предлагаемая Концепция [4] является в той или иной степени развитием Концепции экзафлопных ЭВМ [3] по следующим причинам:

1. Любой из разрабатываемых нейросетевых пакетов прикладных программ разрабатывается, в том числе, и для перспективной реализации гибридной экзафлопной суперЭВМ на базе GPU.

2. Разрабатываемый по предлагаемой Концепции пакет программ распределения нагрузки между процессорами с применением нейросетевых технологий ориентирован в основном и в первую очередь на суперЭВМ, разрабатываемые по Концепции экзафлопных ЭВМ [3].

3. Обе Концепции ставят перед собой одну и ту же цель – создать высокопроизводительные вычислительные системы, по отношению производительности к стоимости (энергопотреблению) значительно превосходящие существующие.

4. Обе Концепции нельзя рассматривать независимо (научно, организационно, финансово) друг от друга. Обе Концепции существуют параллельно друг с другом уже несколько десятилетий. Концепция экзафлопных ЭВМ в первые годы своего существования практически сразу начала трактоваться как «универсальная». Предлагаемая Концепция в течении нескольких десятилетий трактовалась как «специализированная», эффективная для решения узкого класса сложных задач. «Универсальность» предлагаемой Концепции стала активно формироваться в начале 90х годов прошлого столетия и развивается в настоящее время.

### **Литература**

1. **Галушкин А.И.** Транспьютерные системы – начало становления в России ЭВМ с массовым параллелизмом // Информатизация и связь, №1, 2009.

2. Высокопроизводительные вычисления (мировой рынок) // TADVISER, 24.11.2011.

3. Экзафлопные технологии. Концепция по развитию технологий высокопроизводительных вычислений на базе суперЭВМ экзафлопного класса (2012-2020 гг.). 2012г.

4. Экзафлопные нейросетевые технологии. Концепция развития технологий высокопроизводительных вычислений на базе супернейрокомпьютеров (2012-2020 гг.). 2012г.

5. **Галушкин А.И., Симоров С.Н.** Нейросетевые технологии в России (1982-2010) // Из-во «Горячая линия-Телеком», М., 2012г.

6. **Галушкин А.И.** Стратегия развития современных супернейрокомпьютеров на пути к экзафлопным вычислениям // Приложение к журналу «Информационные технологии», №3, 2012.

#### **Сведения об авторах**

**Галушкин Александр Иванович**, Московский физико-технический институт (государственный университет), заместитель заведующего кафедрой «Интеллектуальные информационные системы и технологии». В 1963 окончил Московское высшее техническое училище им. Баумана по кафедре "Системы автоматического управления", в 1966 г. защитил диссертацию на соискание ученой степени кандидата технических наук, в 1969г. получил ученое звание доцента, в 1974г. на Ученом Совете Вычислительного центра АН СССР защитил диссертацию на соискание ученой степени доктора технических наук. Доктор технических наук, профессор. Количество печатных работ и монографий: более 300. Область научных интересов: нейронные сети и нейрокомпьютеры, параллельные вычисления, суперЭВМ, графические процессоры. Email: [neurocomputer@yandex.ru](mailto:neurocomputer@yandex.ru), тел. +7(499)7028197.