

Внедрение новых технологий микроэлектроники в разработки нейрокомпьютеров

Александр Галушкин (Москва)

В статье описываются основные задачи, возникающие при развитии и внедрении новых технологий микроэлектроники в разработки перспективных нейрокомпьютеров, и предлагается методология их решения.

ВВЕДЕНИЕ

Нейросетевые технологии, развиваемые с 1940-х гг., в настоящее время получили активное развитие. Это связано с ростом сложности решаемых задач, что в свою очередь требует отказа от классической фон-Неймановской архитектуры вычислительных систем и перехода к распределённым, в том числе, нейросетевым архитектурам.

Решение сложных задач в нейросетевом логическом базисе нуждается в аппаратной поддержке, которая в настоящее время реализуется на графических процессорах, ПЛИС и, во многих случаях, на цифровых, аналогово-цифровых и аналоговых ИС. В настоящее время известные ограничения цифровых технологий и развитие нанотехнологий делают возможным возврат к аналоговой обработке в рамках высокопараллельных нейросетевых алгоритмов и вычислительных структур.

В течение многих лет кроме стандартной КМОП-технологии для разработки и изготовления нейрочипов имели место активные попытки применения и других технологий:

- мемристорной в сочетании с КМОП-технологией;
- оптической;
- квантовой;
- молекулярной;
- технологии вакуумных схем и др.

Каждая технология накладывала существенные ограничения на структуру аппаратно-реализуемого фрагмента нейронной сети и, как следствие, на алгоритмы настройки коэффициентов и нейросетевые алгоритмы решения задач. Известны российские разработки аналогово-цифровых нейрокомпьютеров 1960–1970 гг., когда весовые коэффициенты реализовывались на потенциометрах, а сумматоры с нели-

нейным преобразованием – на операционных усилителях в составе аналоговых ЭВМ типа МН7 и МН14.

Появление мемристоров позволяет разработать аналоговые нейрочипы с высоким быстродействием. В настоящее время не существует единой концепции создания нейрочипов с использованием мемристоров. Например, компания Hewlett Packard пошла по пути совмещения КМОП-технологии с мемристорными матрицами, а Intel использует совмещение матрицы мемристоров со спиновыми вентилями. Многообразие подходов к конструированию нейрочипов нового поколения свидетельствует о большом потенциале мемристорных элементов.

Принципиальным отличием новых СБИС с мемристорами должна стать частотная модуляция сигнала. Такой подход не только позволит снизить энергопотребление, но и существенно увеличить срок службы и повысить надёжность элементов.

Данная статья посвящена методологии создания нейрокомпьютеров с применением мемристоров, количество которых в СБИС может быть достаточно велико, что, в свою очередь, приведёт к специализации структуры фрагмента аппаратно реализуемой нейронной сети.

В дополнительных материалах к статье (см. на www.soel.ru) представлен перечень работ, необходимых для внедрения новой технологии микроэлектроники (на примере мемристорной) в перспективные нейрокомпьютеры.

Классические подходы к решению задач

Поводом для использования новых технологий микроэлектроники в высокопроизводительных нейрокомпьютерах является актуальная задача или группа задач, которые можно разделить

на два класса: формализуемые и неформализуемые.

Для формализуемых задач совместно с разработкой стандартного алгоритма и его программированием на современной персональной ЭВМ оценивается время решения задачи. Если оно удовлетворяет заказчика (или заказчиков), то потребности во внедрении новой технологии микроэлектроники не возникает. Если время решения задачи не удовлетворяет заказчика, то у разработчика есть два пути:

1. Переход к разработке и программированию алгоритма для кластерной суперЭВМ.
2. Разработка нейросетевого алгоритма решения задачи.

На первом пути анализируются стоимость, габариты, энергопотребление и время решения задачи на кластерной суперЭВМ для различного числа узлов. Если эти характеристики удовлетворяют заказчика, то на этом работа завершается. Если не удовлетворяют, то необходимо разработать и запрограммировать алгоритм решения задачи на суперЭВМ с графическими процессорами. В этом случае, как и в предыдущем, необходимо провести анализ стоимости, габаритов, энергопотребления и времени решения задачи на суперЭВМ с графическими процессорами. Если требования не выполняются, необходимо произвести программирование и реализацию классических алгоритмов решения задачи на базе ПЛИС.

Реализации алгоритмов решения задачи, удовлетворяющие требованиям заказчика, которые не требуют внедрения новых технологий микроэлектроники:

- на ПЭВМ;
- на кластерной суперЭВМ;
- на суперЭВМ с графическими процессорами;
- на базе ПЛИС.

К разработке нейросетевых алгоритмов необходимо переходить, когда:

- время решения задачи на персональной ЭВМ не удовлетворяет заказчика;
- стоимость, габариты, энергопотре-

бление и время решения задачи на кластерных суперЭВМ или суперЭВМ на базе графических процессоров, а также на базе ПЛИС, не удовлетворяют заказчика;

- задача является неформализуемой.

СОЗДАНИЕ И РЕАЛИЗАЦИЯ НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ РЕШЕНИЯ ЗАДАЧ

Разработка нейросетевого алгоритма решения задачи проводится в соответствии с методикой [1], апробированной на многих задачах. После окончания разработки нейросетевого алгоритма решения задачи необходимо выполнить его программирование на ПЭВМ с графическими ускорителями и оценить время решения задачи.

Если время решения задачи удовлетворяет заказчика, то на этом разработка данного варианта нейрокомпьютера заканчивается. В противном случае необходимо провести разработку и программирование нейросетевого алгоритма на кластерной суперЭВМ и т.д. и т.п., включая реализацию на ПЛИС. При отрицательном результате потребуется разработать нейросетевые алгоритмы решения задачи с применением частотно-импульсной модуляции сигналов [2] (Spiking Neural Network).

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ С ЧАСТОТНО-ИМПУЛЬСНОЙ МОДУЛЯЦИЕЙ СИГНАЛОВ

Переход в разработке нейрокомпьютеров к представлению сигналов в виде частотной последовательности коротких импульсов является принципиальным, в первую очередь, с точки зрения энергопотребления. Хорошим примером является нейрочип TrueNorth фирмы IBM [3], который делает нейросетевые технологии (Spiking Neural Network) достаточно самостоятельными и перспективными.

Естественно, начальным этапом этих работ является попытка реализации нейросетевой парадигмы на ПЭВМ, суперЭВМ с графическими ускорителями и вычислительными системами на базе ПЛИС. Вариант реализации на кластерной суперЭВМ здесь пропускается вследствие его неперспективности.

Необходимо отметить, что для варианта нейронных сетей с частотно-импульсной модуляцией сигналов (НСЧИМ) потребуется доработка специфических разделов (теории нейронных сетей, нейроматематики и нейро-

управления), используемых в качестве алгоритмической основы реализации нейрокомпьютеров с применением НСЧИМ.

Первым этапом работ в этом направлении является построение нейросетевого алгоритма решения задачи с применением НСЧИМ. В рамках направления потребуется создание комплекса моделей нейрочипов и нейроморфных ЭВМ на основе НСЧИМ [4], а именно:

- модели входного сигнала нейронных сетей;
- моделей функционалов первичной оптимизации;
- моделей нейронных сетей;
- моделей процессов настройки коэффициентов нейронных сетей в нейроморфных ЭВМ;
- моделей процессов настройки коэффициентов нейронной сети с переменной структурой;
- моделей процессов настройки коэффициентов нейронных сетей в нейроморфных ЭВМ с ограничениями на настраиваемые коэффициенты;
- моделей типовых входных сигналов нейронных сетей;
- планов экспериментов и визуализации результатов для проверки качества работы нейронных сетей;
- моделей решения формализуемых и неформализуемых задач с применением больших нейронных сетей для перспективных нейроморфных ЭВМ;
- моделей распараллеливания нейросетевых алгоритмов в соответствии с реальной структурой физических моделей перспективных нейроморфных ЭВМ.

Указанный комплекс моделей должен быть реализован на суперЭВМ с применением графических процессоров.

Разработанные нейросетевые алгоритмы должны быть запрограммированы и исследованы на ПЭВМ с графическими ускорителями, с анализом стоимости, габаритов, энергопотребления и времени решения задачи для различного числа графических модулей. При удовлетворении требований заказчика будет разработан вариант нейрокомпьютера (ПЭВМ с несколькими графическими платами и нейросетевое (НСЧИМ) программное обеспечение).

Если требования заказчика не удовлетворяются вариантом ПЭВМ с несколькими графическими ускорителями, необходимо перейти к разработке и программированию нейросетевого алгоритма (НСЧИМ) решения задачи

на суперЭВМ и т.д. и т.п., включая ПЛИС. Очевидно, что и эти варианты реализации нейрокомпьютеров с ЧИМ не требуют развития и внедрения новых технологий микроэлектроники.

ПРЕДВАРИТЕЛЬНЫЕ ВЫВОДЫ

Вышерассмотренные варианты реализации нейрокомпьютеров практически не требуют развития и внедрения новых технологий микроэлектроники. Потребность в последних возникает из-за невозможности удовлетворить требования заказчика. Однако необходимо отметить, что:

- отсутствие реализации нейрокомпьютера не может служить основанием для внедрения новых технологий микроэлектроники в перспективные нейрокомпьютеры;
- само по себе развитие и внедрение новых технологий микроэлектроники требует проведения большого объема работ;
- наиболее эффективным вариантом реализации нейрокомпьютера, ориентированного на решение задач в нейросетевом логическом базисе (НСЧИМ) в виде заказной СБИС, в настоящее время является нейрочип TrueNorth [3] фирмы IBM.

РАЗРАБОТКА НЕЙРОЧИПОВ И НЕЙРОМОРФНЫХ ЭВМ С ПРИМЕНЕНИЕМ МЕМРИСТОРОВ

В настоящее время мемристоры являются перспективной технологией для будущих нейрокомпьютеров. Развитие самой технологии изготовления мемристорных матриц является самостоятельным вопросом и не рассматривается в данной статье. Далее описываются проблемы разработки нейрочипов и нейроморфных ЭВМ с применением мемристоров.

Необходимо отметить, что для вариантов нейрочипов и нейроморфных ЭВМ с применением мемристоров необходима доработка специфических разделов:

- теории нейронных сетей, алгоритмы адаптации которых ориентированы на специфические мемристоры структуры нейронных сетей и на учёт динамических свойств мемристоров как элементов, реализующих весовые коэффициенты;
- нейроматематики, как раздела вычислительной математики, связанного с решением задач с учётом указанной модификации теории нейронных сетей, используемой для реше-

ния задач в нейросетевом логическом базисе;

- методов распараллеливания нейросетевых алгоритмов решения задач на специфические «мемристорные», аппаратно реализованные нейросетевые структуры;
- нейроруправления, где для работы в реальном времени в обязательном порядке необходим учёт динамических характеристик мемристоров.

В рамках освоения новой технологии микроэлектроники – мемристоров, разработчикам нейрочипов и нейроморфных ЭВМ совместно с технологами мемристоров необходимо создать модели элементов (мемристоров и мемристорных матриц) с оценкой производительности в операциях будущих нейрочипов и нейроморфных ЭВМ.

Разработчиками нейросетевых (НСЧИМ) алгоритмов решения задач, архитектуры нейрочипов и нейроморфных ЭВМ, специалистами по распараллеливанию нейросетевых алгоритмов должна быть предложена архитектура и схемотехника аналоговой части нейрочипа с применением мемристоров, а совместно с технологами – технология стыковки мемристорных матриц большого размера с аналоговой частью СБИС.

Параллельно должны быть созданы модели указанных аналоговых схем в программных пакетах Cadence или Pspice, а также соответствующие модели аналоговых схем с мемристорными

блоками. Следует также получить оценки производительности данной части нейрочипа в терминах операций нейроморфных ЭВМ.

В результате распараллеливания нейросетевого алгоритма решения задачи должен быть разработан фрагмент нейросетевого алгоритма, реализуемого цифровой частью СБИС, и созданы модели цифровой части в среде Cadence. Интеграция указанных работ с интерфейсом СБИС позволит создать архитектуру, схемотехнику и модели с применением мемристоров в среде Cadence, а также оценить производительность в операциях нейроморфной ЭВМ.

Заключительными этапами работ по созданию нейроморфных ЭВМ будут разработка архитектуры и схемотехники, создание моделей в среде Cadence и оценке производительности:

- нейроплат;
- нейроблоков;
- нейростоек;
- супернейрокомпьютера.

Результаты последних этапов работ будут переданы на производство с учётом библиотек элементов и существующих технологических норм.

ЗАКЛЮЧЕНИЕ

Описанная в статье методология может стать основой планирования работ по внедрению новых технологий микроэлектроники в разработки перспективных нейрокомпьюте-

ров. Любая предлагаемая технология микроэлектроники может иметь более широкую сферу применения, нежели нейрокомпьютеры. Поскольку данная статья посвящена внедрению мемристоров в нейрокомпьютеры, необходимо отметить, что в части массового производства рынок мемристорных систем будет формироваться не за счёт нейрокомпьютеров, а за счёт энергонезависимой памяти, коммутационной среды ПЛИС и других проектов. Это позволяет в значительной степени увеличить допустимые риски разработки нейрокомпьютеров с применением мемристоров.

ЛИТЕРАТУРА

1. Галушкин А.И. О методике решения задач в нейросетевом логическом базисе. Приложение к журналу «Информационные технологии». 2006. №9.
2. Spiking Neural Network. Wikipedia. www.wikipedia.org/wiki/Spiking_neural_network.
3. Akopyan F., Sawada J., Cassidy A., Alvarez-Icaza R., Arthur J., Merolla P., Imam N., Nakamura Y., Datta P., Nam Gi-Joon, Taba B., Beakes M., Brezzo B., Kuang J.B., Manohar R., Risk W. P., Jackson B., Modha D.S. TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2015. Oct. Vol. 34. №10. Pp. 1537–1557.
4. Галушкин А.И. Модели для разработки нейрочипов и нейроморфных ЭВМ. Информатизация и связь. 2015. №4.



Новости мира News of the World Новости мира

Российские физики ускорили синтез наночастиц

Исследователи Сибирского физико-технического института при Томском государственном университете (СФТИ ТГУ) отчитались о новых достижениях в разработке технологий синтеза уникальных наночастиц.

Речь идёт о методе импульсной лазерной абляции в жидкости. Его суть заключается в том, что мишень помещается в жидкость и облучается сфокусированными лазерными импульсами. В результате происходит быстрый разогрев и взрывное испарение материала мишени с поверхности в виде ионизованной плазмы. Этот процесс называется лазерной абляцией. При остывании плазмы в жидкости формируются наночастицы в виде коллоидного раствора.

Российские учёные выявили ряд закономерностей между теплофизическими свойствами материалов мишеней, растворителей, оптическими свойствами наночастиц и особенностями процесса абляции. Результаты исследования позволяют добиваться высокой эффективности синтеза в течение продолжительного времени.

Достижение, в частности, даёт возможность получать сотни литров коллоидных растворов наночастиц или десятки граммов различных нанопорошков в месяц. Такой производительности пока не удалось достичь никому в мире.

Отмечается, что опытные образцы наночастиц используются для исследований в области физики, химии, биологии, медицины, например, для изучения их токсичности и влияния на окружающую среду, при-



менения в качестве антисептиков, биостимуляторов и т.д.

Томский государственный университет