

ПРИМЕНЕНИЕ МИКРОКОНТРОЛЛЕРОВ ДЛЯ РЕАЛИЗАЦИИ НЕЙРОННЫХ СЕТЕЙ

В.А. Руднев

MICROCONTROLLERS APPLICATION FOR NEURAL NETWORK IMPLEMENTATION

V.A. Rudnev

Приведена классификация нейросетевых архитектур по типу элементной базы. Описан ряд процессоров, позволяющих реализовать нейронные сети.

Ключевые слова: нейронная сеть, нейрочип, процессор.

Classification of neural network architecture on the basis of hardware components is given in the article. A range of processors implementing neural networks is described.

Keywords: neural network, neuro chip, processor.

Несмотря на то что нейронные сети как научно-технический объект существуют более полувека, основные способы их практической реализации ограничиваются областью программного моделирования в различных средах визуального (и невизуального) проектирования. Программные реализации преобладают над аппаратными. Кроме того, аппаратные реализации вследствие своей относительной дороговизны до сих пор не имеют повсеместного распространения. Однако их удельный вес в общем количестве мировых технологических разработок неуклонно растет.

Нейросетевые аппаратные решения внедряют в свои продукты такие известные фирмы, как Siemens, Intel, Phillips Research (Нидерланды), 3M Laboratories и многие другие. При этом разработано на удивление много разнообразных аппаратных нейросетевых архитектур. Чтобы не запутаться в этом многообразии, наш небольшой обзор нейросетевых реализаций начнем с классификации, которая представлена на рис. 1.

Примером каскадируемой архитектуры может служить нейрочип NNP фирмы Accurate Automation Corp. Он представляет собой несколько 16-разрядных процессоров, снабженных памятью для хранения весовых коэффициентов. Процессоры связаны локальной внутренней шиной и имеют всего 9 команд. В комплект поставки входит библиотека подпрограмм с реализованными нейросетевыми алгоритмами. Чип поставляется на платах под шины ISA и VME. Производительность составляет от 140 MIPS до 1,4 GIPS, в зависимости

от количества нейрочипов на плате. Их количество может быть от одного до десяти.

Примером RBF-архитектуры может служить чип Ni1000 (разработка фирмы Intel совместно с Nestor), в котором содержится около 3 млн транзисторов, имитирующих 1024 нейрона, быстродействие 33 МГц, 17 GIPS [1]. Выпускается на платах в формате ISA, VME. Кроме того, выпускались ныне уже снятые с производства чипы ZICS18T–ZICS24T, которые также представляли собой нейронные сети на RBF-нейронах.

Сигнальные процессоры характеризует сравнительно низкая стоимость разработки, обширная элементная база с обилием отладочных средств и вспомогательного программного обеспечения, а также высокая производительность, что сделало сигнальные процессоры фирм Texas Instruments и Analog Devices наиболее популярными объектами для построения аппаратных реализаций нейронных сетей. Моделирование нейронных сетей на сигнальных процессорах осуществляется только программным способом, что позволяет в случае необходимости легко модифицировать используемый алгоритм и структуру сети.

Например, использование разработчиками DSP серии ADSP2116x компании Analog Devices оправдывается рядом удачных конструктивных решений, заложенных в архитектуру этих чипов. ADSP2116x используют архитектуру SHARC (Super Harvard ARChitecture), которая предполагает разделение памяти и шин команд и данных, при этом по шине команд могут передаваться как команды, так и дан-

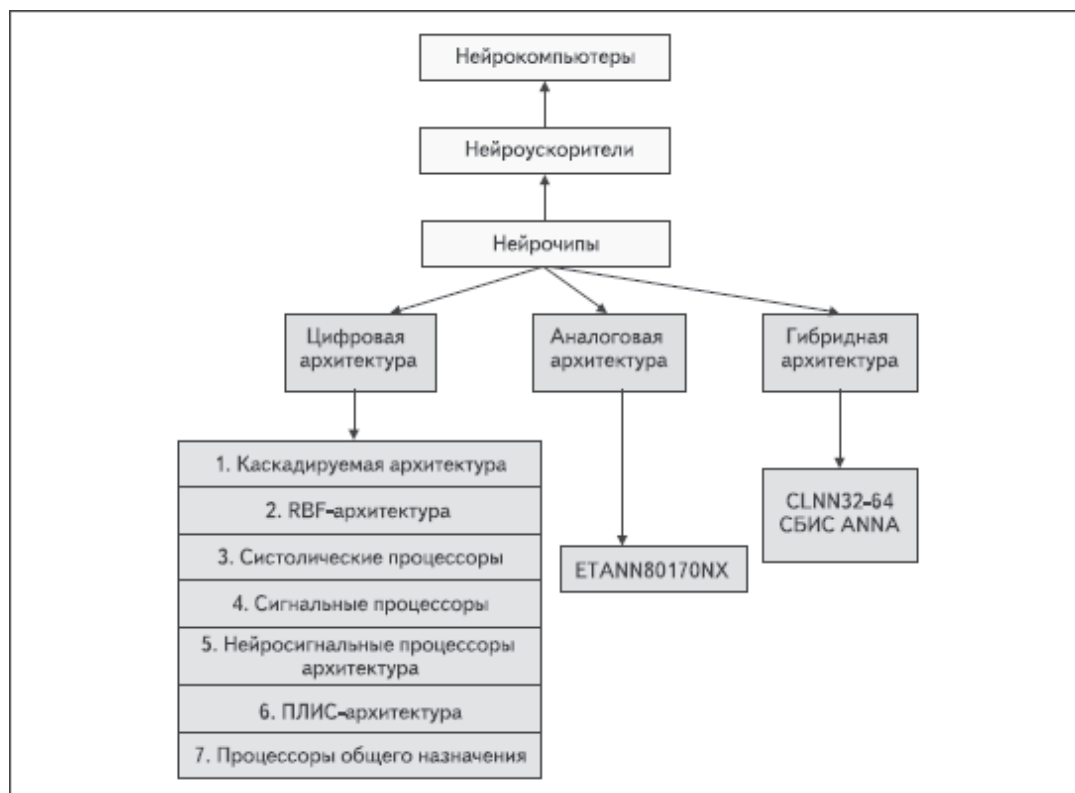


Рис. 1. Классификация нейросетевых архитектур по типу элементной базы

ные. Процессор имеет два генератора адреса данных, что позволяет одновременно генерировать два адреса для выборки двух операндов.

Объем внутренней памяти размером 1 Мбит позволяет реализовать достаточно крупные нейронные сети. Система команд ADSP2116x содержит арифметическую команду умножения с накоплением $C=A \times B + C$, что достаточно удобно для построения циклов вычисления входных значений нейронов. Микропроцессор использует циклические буферы, при обращении к которым происходит автоматическая модификация адресов, что позволяет упростить циклы и увеличить скорость работы сети. Кроме того, ADSP2116x использует два вычислительных блока (обрабатывающих элемента, PEх и PEу) вместо одного.

В стандартном режиме используется только PEх. В режиме SIMD (Single Instruction, Multiple Data) процессор направляет одну и ту же инструкцию в оба процессорных элемента, загружает два отдельных набора данных из памяти в каждый PE, производит параллельные вычисления в обоих PE и синхронизирует результаты в памяти. Таким образом, использование двух параллельно работающих процессорных элементов позволяет в два раза увеличить скорость обработки данных, что весьма актуально при моделировании нейронных сетей, состоящих из множества однотипных элементов [2].

В основе нейросигнальных процессоров лежит то же самое ядро, что и для других распространенных процессоров DSP. Отличие заключа-

ется в добавочных архитектурных решениях, оптимизирующих данные процессоры для разработок на их базе нейронных сетей. В качестве примера приведем наш отечественный нейропроцессор NEUROMATRIX NM6403 фирмы НТЦ «Модуль». Данный процессор состоит из основного 32-разрядного RISC-ядра и добавочного 64-разрядного векторного сопроцессора. NM6403 имеет стандартный интерфейс ввода-вывода, совместимый с портами TMS320C4, может работать с числами переменной разрядности от 1 до 64 бит, его адресное пространство достигает 16 Гбайт. Кроме того, он имеет поддержку векторно-матричных и матрично-матричных операций, а также поддержку двух типов функций насыщения на кристалле. Его производительность по скалярным операциям составляет 120 MIPS (32-разрядные операнды) и от 40 до 11 500 MMAC (миллионов умножений с накоплением в секунду). Желательные могут более подробно ознакомиться с документацией и всеми спецификациями этого процессора на сайте разработчика [3]. Здесь в заключение также отметим, что хотя процессор NM6403 за счет оригинальной архитектуры значительно лучше справляется с матричными операциями, чем другие DSP, однако вся тяжесть программной реализации выбранной нейронной сети по-прежнему лежит на разработчике.

Процессоры общего назначения получили наибольшее распространение в качестве базы для построения нейронных сетей. Этому способствует

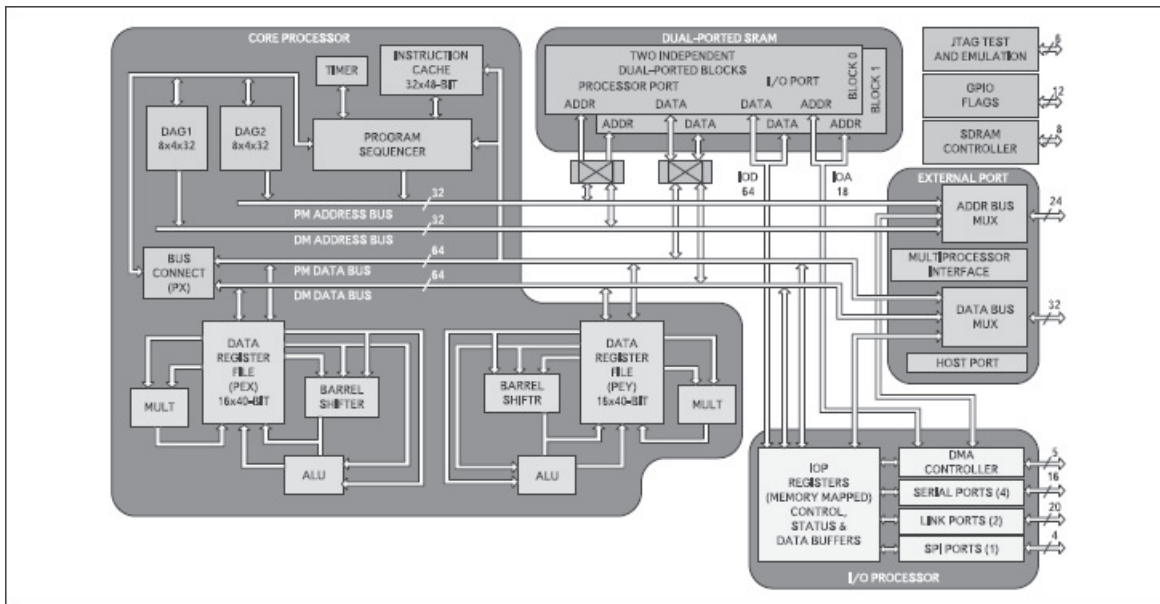


Рис. 2. Архитектура DSP ADSP2116x

широкое распространение персональных компьютеров и впечатляющий рост их производительности. На данный момент создана масса программ и программных библиотек, эмулирующих нейронные сети различных конфигураций. Однако данные нейросетевые реализации осуществляются исключительно на программном уровне, поэтому рассматриваться не будут.

Следует заметить, что при проектировании аппаратной реализации нейросистем учитываются, прежде всего, такие характеристики, как масштабируемость, стоимость разработки, совместимость с предыдущими и будущими версиями [4]. Применительно к рассмотренным категориям чипов это означает, что в случае, когда для разработчика критичны сроки разработки и стоимость, следует обратить внимание на сигнальные процессоры и процессоры для каскадируемых архитектур. Сигнальные процессоры масштабируются довольно плохо, но разработка нейронных сетей на них благодаря развитой технической поддержке со стороны производителей достаточно проста. При этом привязка к аппаратной части DSP не позволит с легкостью перенести нейронную сеть с одной элементной базы на другую. Для нейросигнальных процессоров в целом характерно все то же самое, что справедливо для сигнальных процессоров. От-

личия заключаются в более высокой производительности за счет наличия встроенного векторного сопроцессора и более узкой специализации, в большинстве случаев ориентированной на конкретный вид нейронной сети. Таким образом, общепринятых рекомендаций по выбору определенной элементной базы в настоящее время не существует. Все это в значительной степени определяется личным опытом разработчика, его предпочтениями, а также доступностью на данной территории тех или иных электронных компонентов, и, конечно, в немалой степени зависит от требований самого проекта.

Литература

1. Грибачев, В.П. Настоящее и будущее нейронных сетей / В.П. Грибачев // Компоненты и технологии. – 2006. – № 5. – С. 34–40.
2. Шахнов, В. Нейрокомпьютеры — архитектура и реализация / В. Шахнов, А. Власов, А. Кузнецов. – http://chipnews.gaw.ru/html.cgi/archiv/00_07/stat_36.htm
3. <http://www.module.ru/ruproducts/proc/nm6403.shtml#descr> (сайт фирмы НТЦ «Модуль»).
4. Гриняев, С. Нейронные процессоры Intel / С. Гриняев // Компьютерра. – 2001. – № 38. – http://pgpu.penza.com.ru/sites/intel/MATVEYEV_DM/materials/article_neuropointel.html

Поступила в редакцию 31 мая 2012 г.