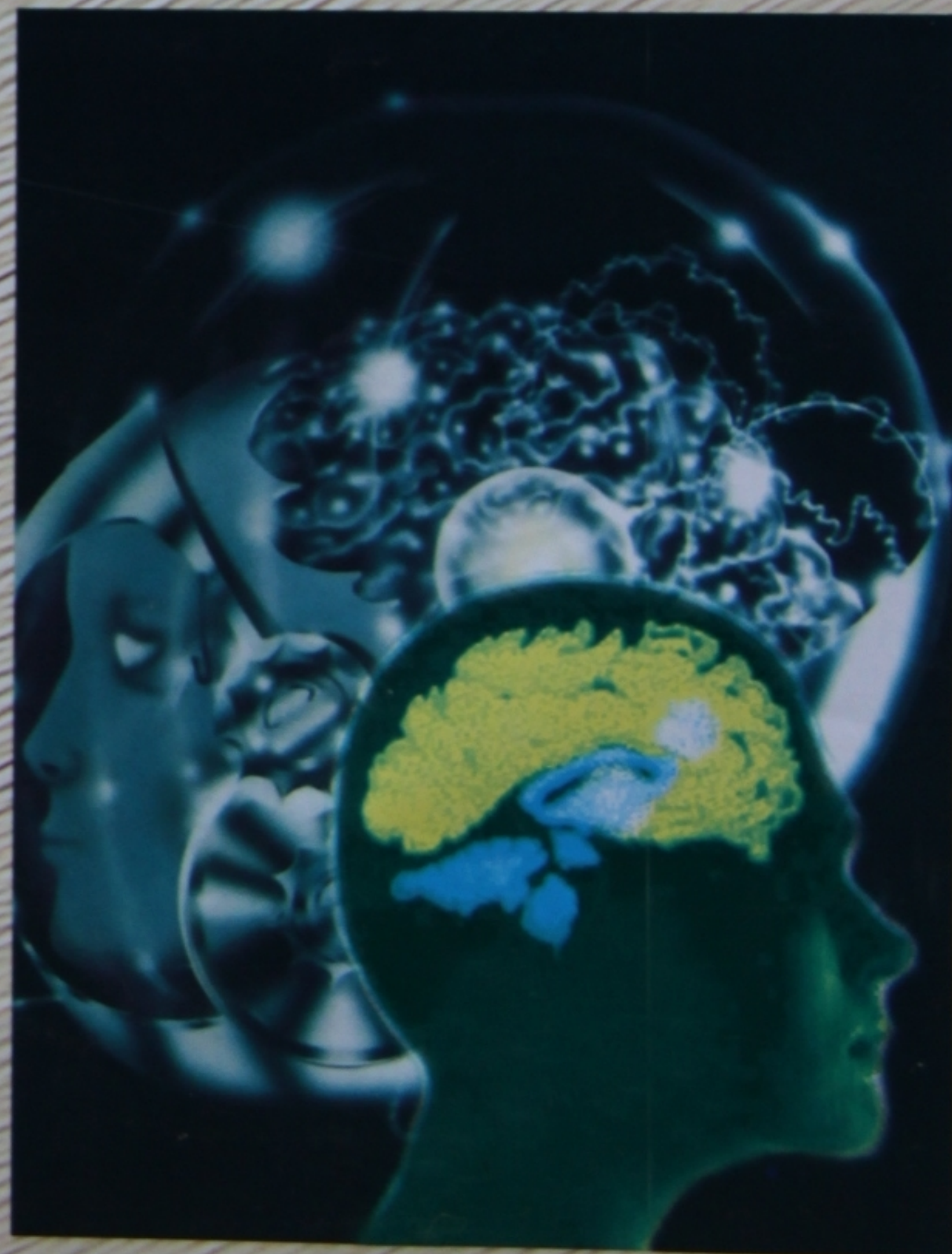


П Р И Л О Ж Е Н И Е К Ж У Р Н А Л У  
«ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ» № 9/2000

**В. А. Шахнов, А.И. Власов, А.С. Кузнецов,  
Ю.А. Поляков**

# Нейрокомпьютеры: архитектура и реализация





А. В. Шахнов, А. И. Власов, А. С. Кузнецов, Ю. А. Поляков

## Нейрокомпьютеры: архитектура и реализация

Главный редактор  
НОРЕНКОВ И. П.

Зам. гл. редактора  
АНТОНОВ Б. И.

Редакционная  
коллегия:

АВДОШИН С. М.  
БАТИЩЕВ Д. И.  
БОЖКО А. Н.  
ГАЛУШКИН А. И.  
ГОРБАТОВ В. А.  
ЗАЙЦЕВА Ж. Н.  
ЗАЛЕЩАНСКИЙ Б. Д.  
ЗАРУБИН В. С.  
ИВАННИКОВ А. Д.  
ИСАЕНКО Р. О.  
КОЛИН К. К.  
КУЛАГИН В. П.  
ЛЬВОВИЧ Я. Е.  
МАЛЬЦЕВ П. П.  
МИХАЙЛОВ Б. М.  
МУХТАРУЛИН В. С.  
НАРИНЬЯНИ А. С.  
НЕЧАЕВ В. В.  
ПАВЛОВ В. В.  
ПУЗАНКОВ Д. В.  
РЯБОВ Г. Г.  
СТЕМПКОВСКИЙ А. Л.  
УСКОВ В. Л.  
ФИЛИМОНОВ Н. Б.

### СОДЕРЖАНИЕ

Введение . . . . .	2
Часть 1. Элементы нейробиологии с позиции аппаратной реализации . . . . .	4
Часть 2. Элементная база нейровычислителей . . . . .	5
2.1. Цифровые сигнальные процессоры (DSP) . . . . .	6
2.2. Программируемые логические интегральные схемы (ПЛИС) . . . . .	10
2.3. Нейрочипы: анализ и сравнительные характеристики . . . . .	11
Часть 3. Аппаратная реализация нейровычислителей . . . . .	16
3.1. Нейроускорители на базе ПЛИС . . . . .	16
3.2. Нейроускорители на базе каскадного соединения сигнальных процессов . . . . .	17
3.3. Нейровычислители на базе нейрочипов . . . . .	21
3.4. Нейрокомпьютеры . . . . .	22
Заключение . . . . .	23
Список литературы . . . . .	24

С 1996 г. аннотации статей размещены на WWW-сервере ГосНИИ информационных технологий и телекоммуникаций Министерства образования РФ и доступны по сети INTERNET.

Адрес сервера: <http://www.informika.ru/text/magaz/it/>

Редакция:

ГРИГОРИН-РЯБОВА Е. В.  
ЛЫСЕНКО А. В.  
(ответственный секретарь)

## Введение

История развития вычислительных систем с массовым параллелизмом насчитывает уже не один десяток лет. Пожалуй, эта одна из небольшого числа областей науки и техники, где отечественные разработки находятся на уровне мировых достижений, а в некоторых случаях и превосходят их. Шли годы, изменялась элементная база и подходы к архитектуре построения современных супервычислителей, появлялись новые направления, к числу которых можно отнести и нейрокомпьютеры.

Что же следует понимать под термином *нейрокомпьютер*? Вопрос достаточно сложный. Нейросетевая тематика, как таковая, является междисциплинарной, ей занимаются как разработчики вычислительных систем и программисты, так и специалисты в области медицины, финансово-экономические работники, химики, физики и т. п. То, что понятно физику, совершенно не принимается медиком, и наоборот, — все это породило многочисленные споры и целые терминологические "войны" по различным направлениям применения всего, где есть приставка *нейро-*.

Общие принципы построения нейронных сетей были заложены в начале второй половины XX века в работах таких ученых, как Д. Хебб, М. Минский, Ф. Розенблат. Первые нейросети состояли из одного слоя искусственных нейронов-персептронов. М. Минским были строго доказаны теоремы, определяющие принципы функционирования нейронных сетей. Несмотря на многочисленные преимущества персептронов (линейность, простота реализации параллельных вычислений, оригинальный алгоритм обучения и т. п.), М. Минским вместе с соавторами было показано, что реализованные на его основе однослойные нейронные сети не способны решить большое число разнообразных задач. Это вызвало некоторое ослабление темпов развития нейросетевых технологий в 60-е годы. В дальнейшем многие ограничения по использованию нейросетей были сняты в связи с разработкой многослойных нейронных сетей, определение которых было впервые введено Ф. Розенблатом: "Под многослойной нейронной сетью понимается такое свойство структуры преобразования, которое осуществляется стандартной разомкнутой нейронной сетью при топологическом, а не символьном описании". Дальнейшее развитие теории нейронных сетей нашла в 70—80 годах в работах Б. Уидроу, Андерсона, Т. Кохонена, С. Гроссберга и др.

Теория нейросетей не внесла революционных новшеств в алгоритмы адаптации и оптимального управления. Самообучающиеся системы известны давно, теория адаптивных регуляторов также хорошо разработана, они широко применяются в технике. Теория нейросетей осваивает разработанные ранее методы и пытается их приспособить для создания все более эффективных нейронных систем. Особую важность использование нейроструктур приобретает с точки зрения производительности ЭВМ. Согласно гипотезе Минского [2—4], реальная производительность типовой параллельной вычислительной системы из  $n$  процессоров растет как  $\log n$  (т. е.

производительность системы из 100 процессоров всего вдвое выше, чем производительность 10-процессорной системы — процессоры дольше ждут своей очереди, чем вычисляют). Однако если использовать для решения задачи нейронную сеть, то параллелизм может быть использован практически полностью, и производительность растет "почти пропорционально"  $n$  [1, 2].

Основные преимущества нейрокомпьютеров связаны с массовым параллелизмом обработки, что обуславливает высокое быстродействие, низкие требования к стабильности и точности параметров элементарных узлов, устойчивость к помехам и разрушениям при большой пространственной размерности системы, причем устойчивые и надежные нейросистемы могут создаваться из низконадежных элементов, имеющих большой разброс параметров.

Приведем некоторые наиболее устоявшиеся определения нейрокомпьютера [1, 2, 4, 6], принятые в конкретных научных областях.

1. **Математическая статистика.** НК — это вычислительная система, автоматически формирующая описание характеристик случайных процессов или их совокупности, имеющих сложные, часто априори неизвестные функции распределения.

2. **Математическая логика.** НК — это вычислительная система, алгоритм работы которой представлен логической сетью элементов частного вида — нейронов, с полным отказом от булевых элементов типа И, ИЛИ, НЕ.

3. **Пороговая логика.** НК — это вычислительная система, алгоритм решения задач в которой представлен в виде сети пороговых элементов с динамически перестраиваемыми коэффициентами и алгоритмами настройки, независимыми от размерности сети пороговых элементов и их входного пространства.

4. **Вычислительная техника.** НК — это вычислительная система с MSIMD-архитектурой, в которой процессорный элемент однородной структуры упрощен до уровня нейрона, резко усложнены связи между элементами и программирование перенесено на изменение весовых коэффициентов связей между вычислительными элементами.

5. **Медицина (нейробиологический подход).** НК — это вычислительная система, представляющая собой модель взаимодействия клеточного ядра, аксонов и дендритов, связанных синаптическими связями (синапсами), т. е. модель биохимических процессов, протекающих в нервных тканях.

6. **Экономика и финансы.** Устоявшегося определения нет, но чаще всего под нейровычислителем понимают систему, обеспечивающую параллельное выполнение "бизнес-транзакций" с элементами "бизнес-логики".

В дальнейшем под НК будем понимать любую вычислительную систему с MSIMD-архитектурой.

Прежде чем перейти к обзору современных нейровычислителей и их элементной базы, остановимся на классификации архитектур вычислительных систем по Б. М. Когану:

- вычислительная система с одним потоком команд и данных (однопроцессорная ЭВМ — SISD) (рис. 1, а);

- вычислительная система с общим потоком команд (SIMD — одиночный поток команд и множественный поток данных) (рис. 1, б);
- вычислительная система с множественным потоком команд и одиночным потоком данных (MISD — конвейерная ЭВМ) (рис. 1, в);
- вычислительная система с множественным потоком команд и данных (MIMD) (рис. 1, г).

Элементарным строительным элементом нейронной сети (НС) является нейрон, который осуществляет взвешенное суммирование поступающих на его вход сигналов. Результат такого суммирования образует промежуточный выходной сигнал, который преобразуется активационной функцией в выходной сигнал нейрона. По аналогии с электронными системами активационную функцию можно считать нелинейной усилительной характеристикой искусственного нейрона, имеющей большой коэффициент усиления для слабых сигналов и с падающим усилением для больших возбуждений. Коэффициент усиления вычисляется как отношение выходного сигнала нейрона к вызвавшему его небольшому приращению взвешенной суммы входных сигналов. Для обеспечения увеличения вычислительной мощности многослойными НС (по сравнению с однослойными) необходимо, чтобы активационная функция между слоями была нелинейной, т. е. с учетом ассоциативности операции умножения матриц любую многослойную нейросеть без нелинейных активационных функций можно свести к эквивалентной однослойной нейросети, которая весьма ограничена по своим вычислительным возможностям. Но вместе с этим наличие нелинейностей на выходе нейрона не может служить определяющим критерием, хорошо известны и успешно работают нейросети и без нелинейных преобразований, получившие название *нейросети на линиях задержки*.

Алгоритмический базис НК обеспечивает теория НС. Нейронная сеть — это сеть с конечным числом слоев из однотипных элементов — аналогов нейронов с различными типами связи между слоями. Среди основных преимуществ НС отмечаются [2]: инвариантность методов синтеза НС к размерности пространства признаков и размерам НС; адекватность современным перспективным технологиям; отказоустойчивость в смысле монотонного, а не катастрофического изменения качества решения задачи в зависимости от числа вышедших из строя элементов.

Решение математических задач в нейросетевом логическом базисе определяют теоретические положения нейроматематики. В [2] выделены следующие этапы решения практически любой задачи в нейросетевом логическом базисе: формирование входного сигнала НС; формирование выходного сигнала НС; формирование желаемого выходного сигнала НС; формирование сигнала ошибки и функционала оптимизации; формирование структуры НС, адекватной выбранной задаче; разработка алгоритма настройки НС, эквивалентного процессу решения задачи в нейросетевом логическом базисе; проведение исследований процесса принятия решения задачи. Все вышеизложенное делает построение современных систем управления с использованием нейросетевого подхода и на основе нейросетевого логического базиса одним из самых перспективных направлений реализации многоканальных и многосвязанных систем управления.

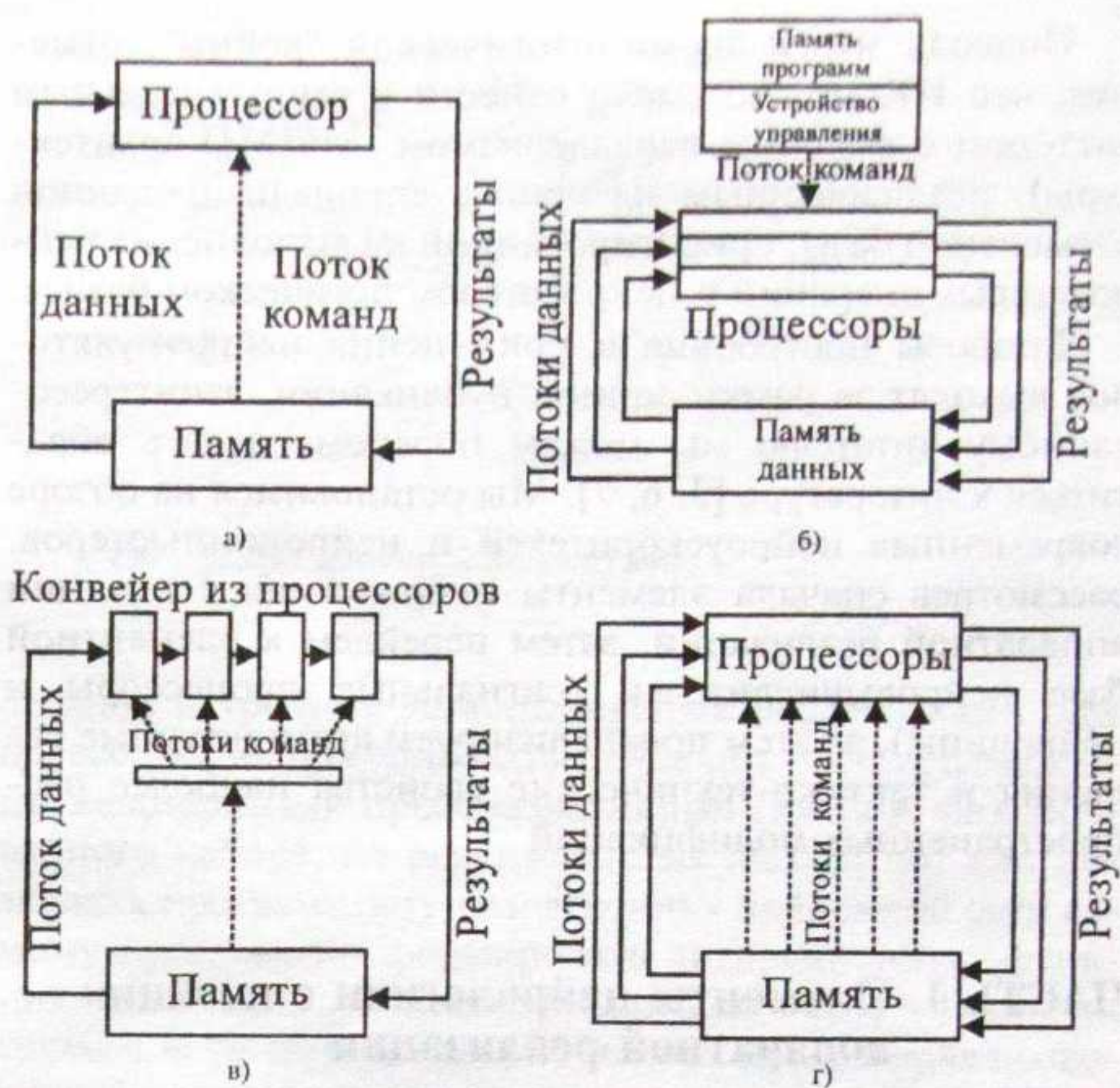


Рис. 1. Архитектуры вычислительных систем

Как было отмечено, НК — это вычислительная система с MSIMD-архитектурой, т. е. с параллельными потоками одинаковых команд и множественным потоком данных. В настоящее время можно выделить три основных направления развития вычислительных систем с массовым параллелизмом (ВСМП):

- ВСМП на базе каскадного соединения универсальных SISD-, SIMD-, MISD-микропроцессоров с элементной базой: универсальные RISC- или CISC-процессоры (Intel, AMD, Sparc, Alpha, Power PC, MIPS и т. п.);
- на базе процессоров с распараллеливанием на аппаратном уровне с элементной базой: DSP-процессоры (TMS, ADSP, Motorola), транспьютеры;
- ВСМП на специализированной элементной базе от специализированных однобитовых процессоров до нейрочипов.

Для каждого из направлений сегодня существуют решения, реализующие те или иные нейросетевые парадигмы. Для большей ясности в дальнейшем будем считать, что нейросетевые системы, реализованные на аппаратных платформах первого направления (пусть и мультипроцессорных) будем относить к *нейроэмуляторам*, т. е. системам, реализующим типовые нейрооперации (взвешенное суммирование и нелинейное преобразование) на программном уровне. Нейросетевые системы, реализованные на аппаратных платформах 2-го и 3-го направлений в виде плат расширения стандартных вычислительных систем (1-го направления) будем называть *нейроускорителями* (взвешенное суммирование, как правило, реализуется аппаратно, например на основе трансверсальных фильтров, а нелинейные преобразования — программно). Системы, реализованные на аппаратной платформе 3-го направления в виде функционально законченных вычислительных устройств следует относить к *нейрокомпьютерам* (все операции выполняются в нейросетевом логическом базисе).

Подводя итоги терминологической "войны", отметим, что НК можно смело отнести к вычислительным системам с высоким параллелизмом (MSIMD-архитектуры), реализованным на основе специализированной элементной базы, ориентированной на выполнение нейросетевых операций в нейросетевом логическом базисе.

Вопросы построения и применения нейроэмуляторов выходят за рамки данной публикации, заинтересованному читателю мы можем порекомендовать обратиться к литературе [3, 6, 7]. Мы остановимся на обзоре современных нейроускорителей и нейрокомпьютеров, рассмотрев сначала элементы нейробиологии с позиции аппаратной реализации, затем перейдем к элементной базе нейровычислителей (сигнальные процессоры и нейрочипы), а затем проанализируем архитектурные решения и тактико-технические свойства наиболее распространенных модификаций.

## ЧАСТЬ 1. Элементы нейробиологии с позиции аппаратной реализации

Основные нейросетевые парадигмы разработаны несколько десятилетий назад, по их исследованию опубликовано огромное число работ, с обзорами которых можно познакомиться в [1—4]. Мы лишь, для лучшего понимания в дальнейшем архитектурно-схемотехнических решений нейровычислительных систем, остановимся на наиболее важных элементах нейробиологии с позиции аппаратной реализации.

Одним из основных достоинств нейровычислителя является то, что его основу составляют относительно простые, чаще всего однотипные, элементы (ячейки), имитирующие работу нейронов мозга — "нейроны". Каждый нейрон характеризуется своим текущим состоянием по аналогии с нервными клетками головного мозга, которые могут быть возбуждены или заторможены. Он имеет группу *синапсов* — однонаправленных входных связей, соединенных с выходами других нейронов, а также *аксон* — выходную связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов. Общий вид нейрона приведен на рис. 2.

Каждый синапс характеризуется величиной синаптической связи, или ее весом  $w_i$ , который по физическому смыслу эквивалентен электрической проводимости. Текущее состояние нейрона определяется как взвешенная сумма его входов:

$$s = \sum_{i=1}^n x_i w_i.$$

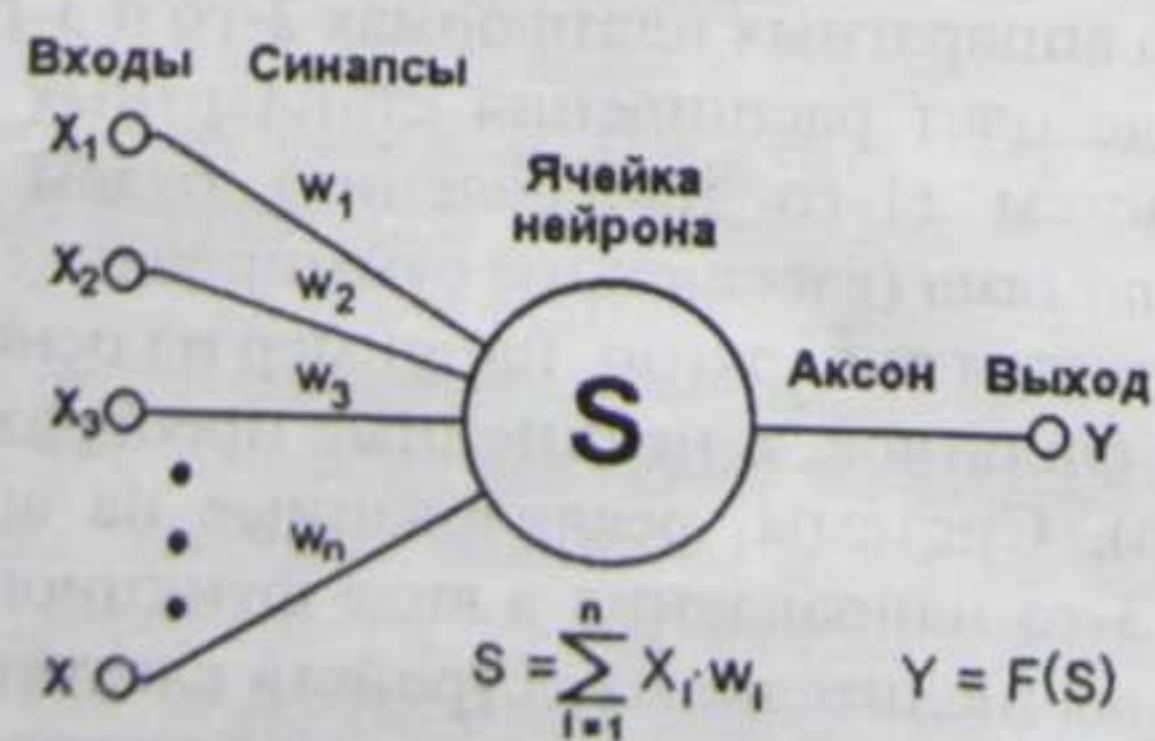


Рис. 2. Общий вид нейрона

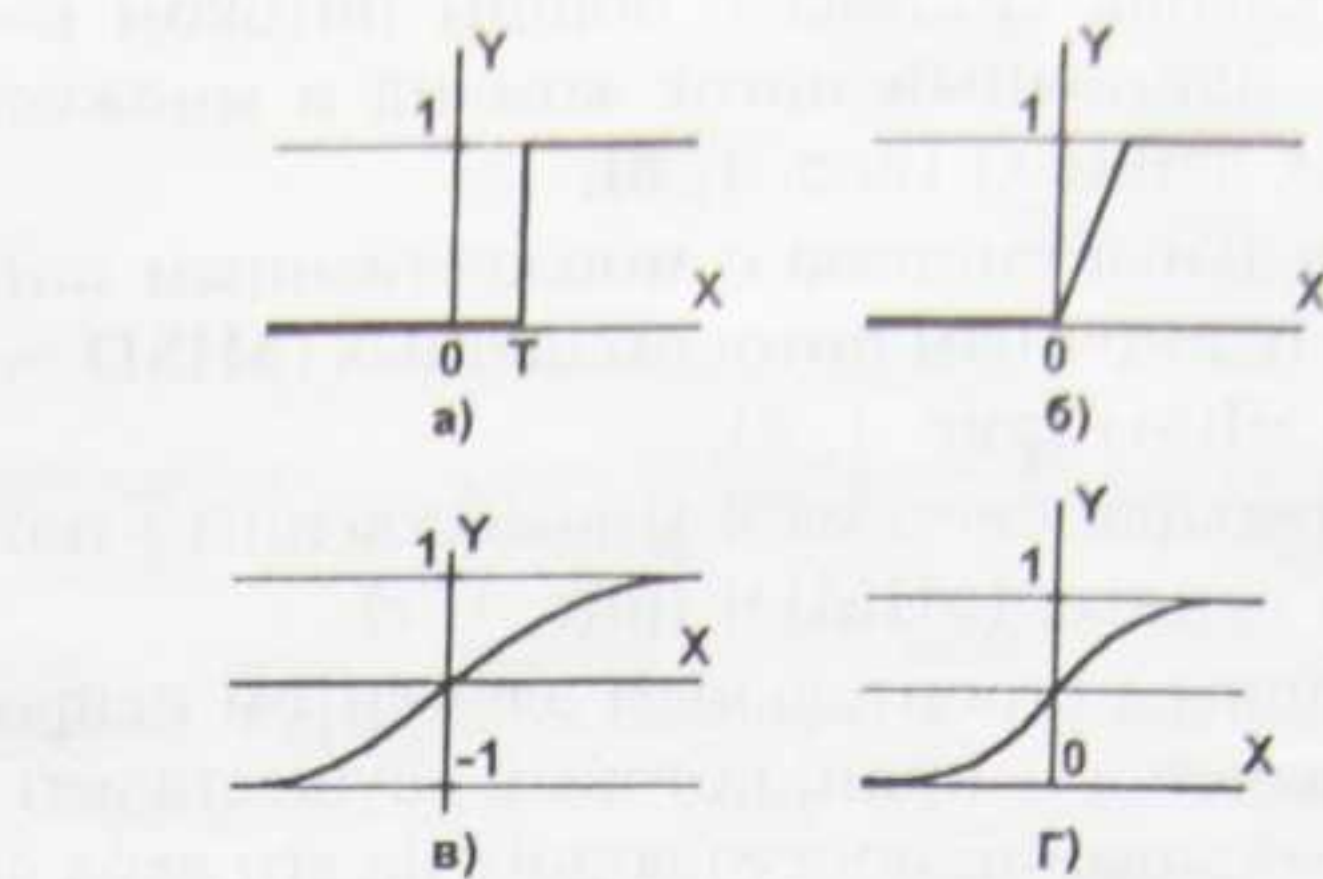


Рис. 3. Активационная функция:

*a* — единичная пороговая; *б* — линейный порог (гистерезис); *в* — сигмоид — гиперболический тангенс; *г* — логический сигмоид

Выход нейрона есть функция его состояния  $y = f(s)$ , которая называется *активационной* и может иметь различный вид (некоторые из типовых активационных функций представлены на рис. 3). Одной из наиболее распространенных является нелинейная функция с насыщением — так называемая логистическая функция, или сигмоид (т. е. функция S-образного вида):

$$f(x) = \frac{1}{1 + e^{-\alpha x}}.$$

При уменьшении  $\alpha$  сигмоид становится более пологим, в пределе при  $\alpha = 0$  вырождаясь в горизонтальную линию на уровне 0,5, при увеличении  $\alpha$  сигмоид приближается по внешнему виду к функции единичного скачка с порогом  $T$  в точке  $x = 0$ . Из выражения для сигмоида очевидно, что выходное значение нейрона лежит в диапазоне  $[0, 1]$ . Одно из ценных свойств сигмоидной функции — простое выражение для ее производной:

$$f'(x) = \alpha \cdot f(x) \cdot (1 - f(x)).$$

Параллелизм обработки достигается путем объединения большого числа нейронов в слои и соединения определенным образом различных нейронов между собой. В качестве примера простейшей НС приведем трехнейронный перцептрон (рис. 4), нейроны которого имеют активационную функцию в виде единичной пороговой функции, работа которого подробно рассмотрена в литературе [2—4]. На  $n$  входов поступают некие сигналы, проходящие по синапсам на три нейрона, которые об-

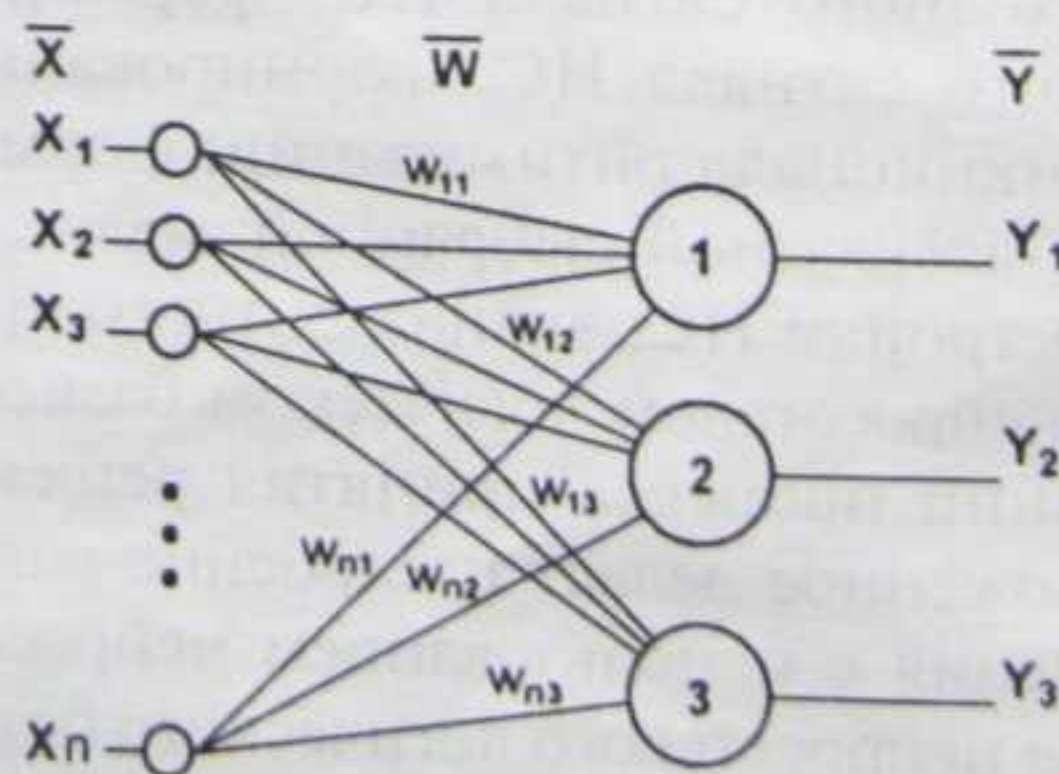


Рис. 4. Однослойный перцептрон

разуют единственный слой этой НС и выдают три выходных сигнала:

$$y_j = f \left[ \sum_{i=1}^n x_i w_{ij} \right], \quad j = 1 \dots 3.$$

Очевидно, что все весовые коэффициенты синапсов одного слоя нейронов можно свести в матрицу  $W$ , в которой каждый элемент  $w_{ij}$  задает величину  $i$ -й синаптической связи  $j$ -го нейрона. Таким образом, процесс, происходящий в НС, может быть записан в матричной форме:  $Y = F(XW)$ , где  $X$  и  $Y$  — соответственно входной и выходной сигнальные векторы;  $F(V)$  — активационная функция, применяемая поэлементно к компонентам вектора  $V$ . Теоретически число нейронов в каждом слое может быть произвольным.

Для того чтобы нейронная сеть работала, ее надо обучить. От качества обучения зависит способность сети решать поставленные перед ней проблемы. На этапе обучения кроме параметра качества подбора весовых коэффициентов важную роль играет время обучения. Как правило, эти два параметра связаны обратной зависимостью и их приходится выбирать на основе компромисса. Обучение НС может вестись с учителем или без него. В первом случае сети предъявляются значения как входных, так и желательных выходных сигналов, и она по некоторому внутреннему алгоритму подстраивает веса своих синаптических связей. Во втором случае выходы НС формируются самостоятельно, а веса изменяются по алгоритму, учитывающему только входные и производные от них сигналы.

Рассматривая классификацию НС можно выделить: бинарные (цифровые) и аналоговые НС, предварительно обученные (неадаптивные) и самообучающиеся (адаптивные) нейронные сети, что крайне важно при их аппаратной реализации. Бинарные оперируют с двоичными сигналами, и выход каждого нейрона может принимать только два значения: логический ноль ("заторможенное" состояние) и логическая единица ("возбужденное" состояние). К этому классу сетей относится и рассмотренный выше трехнейронный персептрон, так как выходы его нейронов, формируемые функцией единичного скачка, равны либо 0, либо 1. В аналоговых сетях выходные значения нейронов могут принимать непрерывные значения, что могло бы иметь место после замены активационной функции нейронов персептрона на сигмоид.

Сети также можно классифицировать по топологии (числу слоев и связей между ними). На рис. 5 представлен двухслойный персептрон, полученный из персептрона рис. 4 путем добавления второго слоя, состоящего из двух нейронов. При этом нелинейность активационной функции имеет конкретный смысл: если бы она не обладала данным свойством или не входила в алгоритм работы каждого нейрона, результат функционирования любой  $p$ -слойной НС с весовыми матрицами  $W^{(i)}$ ,  $i = 1, 2, \dots, p$ , для каждого слоя  $i$  сводился бы к перемножению входного вектора сигналов  $X$  на матрицу  $W^{(\Sigma)} = W^{(1)} \cdot W^{(2)} \cdot \dots \cdot W^{(p)}$ , т. е. фактически такая  $p$ -слойная НС эквивалентна однослойной НС с весовой матрицей единственного слоя  $W^{(\Sigma)}$ :  $Y = XW^{(\Sigma)}$ .

Вопросы, стоящие перед разработчиками нейросетевой элементной базы и нейровычислителей, во многом

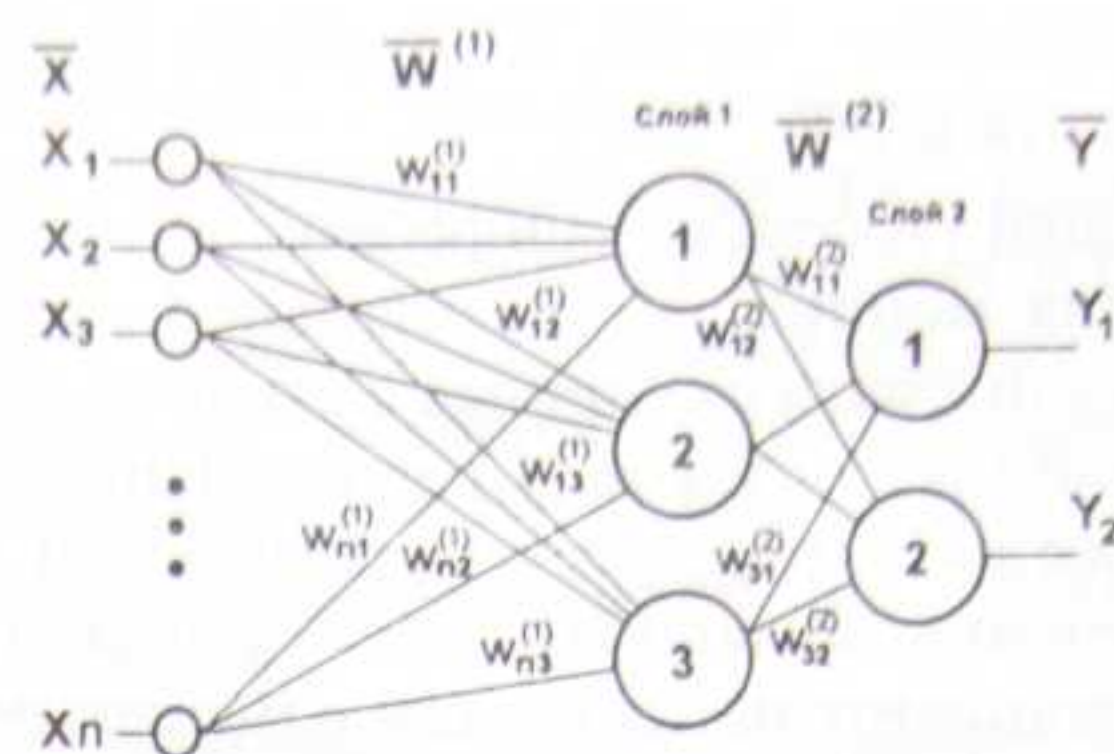


Рис. 5. Двухслойный персептрон

сложны и требуют дополнительных исследований: как целесообразней реализовать нейрочип — со встроенными нелинейными преобразованиями (пусть и фиксированного набора, но реализованных аппаратно) или позволить программисту-разработчику нейронной сети самому программно формировать активационную функцию (размещая соответствующий код в ПЗУ). Стоит ли гнаться за универсальностью нейрочипа, которая позволила бы реализовывать любые топологии и парадигмы, или следует ориентироваться на выпуск специализированных предметно ориентированных нейрочипов, использовать ли базовые матричные кристаллы (БМК) и ПЛИС — все эти вопросы требуют ответа (по крайней мере, обсуждения). Мы предлагаем всем заинтересованным лицам продолжить данное обсуждение в on-line форуме на сервере "Новости с Российского рынка нейрокомпьютеров" (<http://neurnews.iu4.bmstu.ru>).

*Подводя итоги первого раздела, заметим, что эффективное применение НК характерно для случаев, требующих резкого сокращения времени обработки при решении пространственных задач повышенной размерности, которые можно найти практически в любой области деятельности. В следующих разделах мы подробнее остановимся на элементной базе нейровычислителей и архитектуре их построения.*

## ЧАСТЬ 2. Элементная база нейровычислителей

Элементной базой нейровычислительных систем 2-го и 3-го направлений являются заказные кристаллы (ASIC), встраиваемые микроконтроллеры ( $\mu C$ ), процессоры общего назначения (GPP), программируемая логика (FPGA — ПЛИС), транспьютеры, цифровые сигнальные процессоры (DSP) и нейрочипы [1]. Использование как тех, так и других позволяет сегодня реализовывать нейровычислители, функционирующие в реальном масштабе времени, однако наибольшее применение при реализации нейровычислителей нашли ПЛИС, DSP и, конечно, нейрочипы.

Как отмечено в [2], транспьютеры (T414, T800, T9000) и, в частности, транспьютероподобные элементы играют важную роль при построении вычислительных систем с массовым параллелизмом, а их применение постепенно сдвигается в сторону коммутационных систем и сетей ЭВМ, хотя еще остаются примеры реализации на них слоев некоторых ЭВМ с массовым параллелизмом в виде решеток процессорных элементов. Подробнее о транспьютерных системах и их применении при по-

строении вычислительных систем с массовым параллелизмом можно узнать в [3].

DSP (Digital Signal Processor — цифровой сигнальный процессор), обладая мощной вычислительной структурой, позволяют реализовать различные алгоритмы обработки информационных потоков. Сравнительно невысокая цена, а также развитые средства разработки программного обеспечения позволяют легко применять их при построении вычислительных систем с массовым параллелизмом.

Стремительный переход современных систем управления на цифровые стандарты привел к необходимости обрабатывать с высокой скоростью достаточно большие объемы информации. Сложная обработка и фильтрация сигналов (например, распаковка сжатых аудио- и видеоданных, маршрутизация информационных потоков и т. п.) требуют применения достаточно производительных вычислительных систем. Подобные системы могут быть реализованы на различной элементной базе, но наибольшее распространение получили устройства с применением **цифровых сигнальных процессоров и ПЛИС**.

Программируемая логика способна работать на более высоких частотах, но поскольку управление реализовано аппаратно, то изменение алгоритмов работы требует перепрограммирования ИС. Низкая тактовая частота DSP пока ограничивает максимальную частоту обрабатываемого аналогового сигнала до уровня  $\approx 20$  МГц, но программное управление позволяет достаточно легко изменять не только режимы обработки, но и функции, выполняемые DSP. Помимо обработки и фильтрации данных DSP могут осуществлять маршрутизацию цифровых потоков, вырабатывать управляющие сигналы и даже формировать сигналы системных шин ISA, PCI и др.

Оценивать быстродействие тех или иных устройств на основе DSP и ПЛИС принято по времени выполнения типовых операций цифровой обработки сигналов (фильтр Собеля, БПФ, преобразование Уолша—Адамара и др.). Однако оценки производительности нейровычислителей используют другие показатели:

- **CUPS** (connections update per second) — число измененных значений весов в секунду (оценивает скорость обучения);
- **CPS** (connections per second) — число соединений (умножений с накоплением) в секунду (оценивает производительность);
- **CPSPW** =  $CPS/Nw$ , где  $Nw$  — число синапсов в нейроне;
- **CPPS** — число соединений примитивов в секунду,  $CPPS = CPS * Bw * Bs$ , где  $Bw$ ,  $Bs$  — разрядность весов и синапсов;
- **ММАС** — миллионы умножений с накоплением в секунду.

Особенностью использования DSP и ПЛИС в качестве элементной базы нейровычислителей является то, что ориентация в выполнении нейросетевых операций обуславливает, с одной стороны, повышение скоростей обмена между памятью и параллельными арифметическими устройствами, а с другой стороны — уменьшение времени весового суммирования (умножения и накопления) за счет применения фиксированного набора команд типа регистр—регистр [1—11].

## 2.1. Цифровые сигнальные процессоры (DSP)

DSP вот уже на протяжении нескольких десятилетий являются элементной базой для построения как нейроускорителей, так и контура логики общесистемного управления НК. Для реализации нейроускорителей могут использоваться практически любые DSP, все зависит лишь от фантазии и возможностей. Мы проанализируем микропроцессоры лишь трех ведущих производителей — *Analog Devices*, *Motorola* и *Texac Instruments*, с позиций построения на их основе нейровычислительных систем.

Выбор того или иного процессора — многокритериальная задача, однако, следует отметить предпочтительность процессоров *Analog Devices* [7] для приложений, требующих выполнения больших объемов математических вычислений (таких, как цифровая фильтрация сигнала, вычисление корреляционных функций и т. п.), поскольку их производительность на подобных задачах выше, чем у процессоров компаний *Motorola* и *Texac Instruments*. В то же время для задач, требующих выполнения интенсивного обмена с внешними устройствами (многопроцессорные системы, различного рода контроллеры), предпочтительнее использовать процессоры *Texac Instruments* [8], обладающие высокоскоростными интерфейсными подсистемами. Компания *Motorola* является лидером по объему производства сигнальных микропроцессоров, большую часть которых составляют дешевые и достаточно производительные 16- и 24-разрядные микропроцессоры с фиксированной точкой. Расширенные коммуникационные возможности, наличие достаточных объемов внутрикристалльной памяти для данных и программ, возможность защиты программы от несанкционированного доступа, поддержка режима энергосбережения делают эти микропроцессоры привлекательными для использования не только в качестве специализированных вычислителей, но и в качестве контроллеров, в бытовых электронных приборах, в системах адаптивной фильтрации и т. д.

Большая производительность, требуемая при обработке сигналов в реальном времени, побудила *Texac Instruments* и *Analog Devices* выпустить транспьютероподобные семейства микропроцессоров TMS320C4x и ADSP2106x, ориентированные на использование в мультипроцессорных системах. На этом фоне первый российский сигнальный процессор (нейросигнальный процессор) фирмы "Модуль" — "Neuro Matrix" [6] — выглядит весьма достойно среди DSP, работающих в формате с фиксированной точкой. При тактовой частоте 50 МГц "Neuro Matrix" практически не уступает по производительности изделиям мировых лидеров, а по некоторым задачам даже превосходит их.

При создании нейровычислительных систем на базе сигнальных процессоров необходимо помнить, что DSP обладают высокой степенью специализации. В них широко используются методы сокращения длительности командного цикла, характерные для универсальных RISC-процессоров, такие как конвейеризация на уровне отдельных микроинструкций и инструкций, размещение операндов большинства команд в регистрах, использование теневых регистров для сохранения состояния вычислений при переключении контекста, разделение шин команд и данных (*Гарвардская архитектура*). В то же время для сигнальных процессоров характерным является наличие аппаратного умножителя, позволяющего

выполнять умножение как минимум двух чисел за один командный такт.

Другой особенностью сигнальных процессоров является включение в систему команд таких операций, как умножение с накоплением  $C = A \times B + C$  с указанным в команде числом выполнений в цикле и с правилом изменения индексов используемых элементов массивов  $A$  и  $B$ , т. е. уже реализованы прообразы разовых нейроопераций — взвешенное суммирование с накоплением, разнообразные битовые операции. В сигнальных процессорах реализуется аппаратная поддержка программных циклов, кольцевых буферов, один или несколько операндов извлекаются из памяти в цикле исполнения команды.

Реализация однократного умножения и команд, использующих в качестве операндов содержимое ячеек памяти, обуславливает сравнительно низкие тактовые частоты работы сигнальных процессоров. Специализация не позволяет поднимать производительность за счет быстрого выполнения коротких команд типа  $R, R \rightarrow R$ , как это делается в универсальных процессорах. Этим команд нет в программах цифровой обработки сигналов.

Сигнальные процессоры различных компаний-производителей образуют два класса, существенно различающиеся по цене: более дешевые микропроцессоры для обработки данных в формате с фиксированной точкой и более дорогие микропроцессоры, аппаратно поддерживающие операции над данными в формате с плавающей точкой.

Типичные DSP-операции требуют выполнения множества простых сложений и умножений и заключаются в следующем:

- выборка двух операндов;
- выполнение сложения или умножения (обычно и то, и другое);
- сохранение результата или удержание его до повторения.

Для выборки двух операндов за один командный цикл необходимо осуществить два доступа к памяти одновременно. Но в действительности кроме выборки двух операндов требуется еще сохранить результат и прочесть саму инструкцию, поэтому число доступов в память за один командный цикл будет больше двух и, следовательно, DSP-процессоры поддерживают множественный доступ к памяти за один и тот же командный цикл. Однако невозможно осуществить доступ к двум различным адресам в памяти одновременно, используя для этого одну шину памяти. Существует два вида архитектур DSP-процессоров, позволяющих реализовать механизм множественного доступа к памяти: Гарвардская архитектура; модифицированная архитектура фон Неймана.

**Гарвардская архитектура** имеет две физически разделенные шины данных. Это позволяет осуществить два доступа к памяти одновременно: традиционная Гарвардская архитектура выделяет одну шину для выборки инструкций (шина адреса), а другую — для выборки операндов (шина данных). Но для выполнения DSP-операций этого недостаточно, так как в основном все они используют по два операнда. Поэтому Гарвардская архитектура применительно к цифровой обработке сигналов использует шину адреса и для доступа к данным. Важно отметить, что часто необходимо выполнить выборку трех компонентов — инструкции с двумя операндами, на что собственно Гарвардская архитектура неспо-

собна. В таком случае данная архитектура включает в себя кэш-память. Она может быть использована для хранения тех инструкций, которые будут использоваться вновь. При применении кэш-памяти шина адреса и шина данных остаются свободными, что делает возможным выборку двух операндов. Такое расширение — Гарвардская архитектура плюс кэш — называют **расширенной Гарвардской архитектурой**, или **SHARC** (Super Harvard ARChitecture).

Гарвардская архитектура требует наличия двух шин памяти. Это значительно повышает стоимость производства чипа. Так, например, DSP-процессор, работающий с 32-битными словами и в 32-битном адресном пространстве требует наличия по крайней мере 64 выводов для каждой шины памяти, а в сумме получается 128 выводов. Это приводит к увеличению размеров чипа и к трудностям при проектировании схемы.

**Архитектура фон Неймана** использует только одну шину памяти и обладает рядом положительных черт: более дешевая; требует меньшего числа выводов шины; является более простой в использовании, так как программист может размещать и команды и данные в любом месте свободной памяти.

С точки зрения реализации нейроускорителей мы остановимся только на некоторых наиболее ярких представителях DSP, в основном относящихся к классу транспьютероподобных DSP, работающих в формате с плавающей точкой.

### DSP фирмы Analog Devices

Реализация нейровычислителей высокой пространственной размерности требует все более производительной элементной базы. Для преодоления возникающих трудностей разработчики используют два возможных подхода: улучшение характеристик уже имеющихся процессоров; увеличение производительности путем разработки новых архитектур. Первый способ ограничен увеличением производительности в 5—8 раз. Второй способ предполагает разработку архитектур, которые были бы наиболее удобны в конечном приложении и оптимизированы для конкретного языка программирования.

Компания *Analog Devices* [7] ведет разработки в обоих направлениях. Так, ядро первого 32-разрядного процессора ADSP-21020 производительностью 30 MFLOPS было усовершенствовано, что привело к созданию нового процессора ADSP-21065L с максимальной производительностью 198 MFLOPS, что соответствует ускорению в 6,6 раз. Дальнейшее увеличение производительности и оптимизация архитектуры существующих процессоров позволили разработать семейство новых сигнальных микропроцессоров ADSP-2116x с тактовой частотой  $\approx 100$  МГц и средней производительностью 600 MFLOPS.

Среди основных особенностей DSP семейства ADSP-2116x можно отметить:

- **быстрые и гибкие модули арифметики.** Все команды выполняются за один такт. Набор команд микропроцессора наряду с традиционными арифметическими операциями включает такие, как  $1/x$ ,  $1/R(x)$ , команды сдвига, циклического сдвига, комбинации операций сложения/вычитания с умножением;
- **независимые потоки данных  $v(\text{из})$  вычислительные( $x$ ) модули( $ей$ ).** За один такт процессор может одновременно считать (записать) два операнда в регистровый



файл, загрузить два операнда в АЛУ, принять два операнда в умножитель, АЛУ и умножитель могут вырабатывать два результата (или три, если АЛУ выполняет операцию совместно со сложением/вычитанием). 48-битовое командное слово позволяет задавать в одной инструкции параллельное выполнение арифметических операций и обмен данными;

- *повышенная точность и расширенный динамический диапазон выполняемых операций.* Все представители микропроцессорного семейства оперируют с данными в 32-битовом формате с плавающей точкой, 32-битовыми целочисленными данными (в дополнительном коде и беззнаковыми) и 40-битовыми данными расширенной точности с плавающей точкой. Повышенная точность вычислений достигается благодаря уменьшению ошибки округления результата в вычислительных модулях. Аккумулятор для 32-разрядных данных с фиксированной точкой имеет 80 разрядов;
- *наличие двух генераторов адреса.* Генераторы адреса обеспечивают пред- или постформирование прямого или косвенного адреса данных, выполняют над адресами модульные и бит-реверсные операции;
- *эффективные средства формирования последовательности команд и механизм организации программных циклов.* Инициализация, возврат на начало и выход из программного цикла выполняется за один процессорный цикл для уровня вложенности до 6. Процессор аппаратно поддерживает выполнение команд перехода и перехода с задержкой.

Процессоры ADSP-216xx содержат высокопроизводительный кэш команд. Кэш работает избирательно: кэшируются только те команды, выборка которых конфликтует с выборкой данных из памяти программ (Program Memory, PM).

**TigerSHARC.** Современные требования рынка обусловили появление новой архитектуры — TigerSHARC (рис. 6, см. вторую сторону обложки), которая для получения высокого уровня производительности объединяет в себе множество особенностей ранее разработанных архитектур. Новый процессор объединил в себе достоинства, присущие существующим DSP-технологиям, такие как быстрота и определенность времени выполнения команд, распознавание быстрых прерываний и высокая скорость обмена данными с периферийными устройствами.

Для достижения наивысшей производительности ядра использованы особенности RISC-архитектуры, такие как структура хранения и обмена данными и командами, высоко конвейеризированный генератор адресов команд с возможностью предсказания переходов и объемным блокирующим файлом регистров. Также было решено взять в расчет особенности VLIW-архитектуры для оптимизации построения команд. Полученная архитектура TigerSHARC достигает производительности 1,5 GFLOPS при 32-разрядных вычислениях с плавающей точкой и 6 GFLOPS при 16-разрядных вычислениях с фиксированной точкой. Тактовая частота составляет 250 МГц.

**Архитектура TigerSHARC.** По данным компании Analog Devices [7] большинство нейросетевых реализаций на базе DSP схожи в использовании команд, но различаются набором данных. В многоканальных приложениях или в тех случаях, когда данные располагаются в виде прямоугольной матрицы, производительность может быть удвоена путем добавления второго набора математических

модулей. Процессоры, содержащие второй вычислительный модуль, управляются так же, как и небольшие SIMD-архитектуры. TigerSHARC позволяет использовать единственную команду для обработки данных в обоих вычислительных модулях — это уже элементы нейрочипа!

Более того, второй модуль может функционировать независимо от первого, для этой модели используются частично-множественные пути к данным. Для правильного выполнения команды в ней должно быть выделено дополнительное место, что приводит к очень длинным командным словам (VLIW — Very Long Instruction Word). Их использование сопровождается в большинстве случаев к быстрым заполнением небольшого объема внутрикристалльной памяти командами NOP (no operation), которые предназначены для тех устройств, которые не будут задействованы в текущем цикле. Размещение команды NOP в коде в существующих VLIW-разработках — большой недостаток, который был устранен делением длинных слов на более мелкие, параллельно поступающие к каждому устройству. Обработка множества команд независимыми устройствами одновременно является главной особенностью суперскалярной процессорной архитектуры.

**Вычислительные устройства TigerSHARC.** Сердцем любого процессора является его вычислительная подсистема. Каждый из двух вычислительных модулей — Processing Element X (PEX) и Y (PEY) содержит 32-входовый (по 32 бита в каждом входе) блокировочный регистровый файл. При каждом вычислении, выполняемом АЛУ, МАС или Shifter, данные будут выбраны из этого регистрового файла, а затем в него будут помещены результаты вычислений, что является главной особенностью реализации считывания/записи. Использование большого числа регистров для хранения данных упрощает использование высокоуровневых языков программирования. Для достижения высокой внутренней пропускной способности каждый регистровый файл соединен с тремя 128-битовыми шинами посредством двух 128-битных шин. Обе шины могут использоваться одновременно для выполнения операций чтения из памяти и одна шина может быть использована для записи в память. Такая структура шин определяет типичные математические инструкции, требующие выполнения двух операций чтения данных и закачивающиеся записью результата в память.

Сравнительно большой объем внутрикристалльной памяти разделен на три независимых блока одинаковой величины. Каждый блок имеет ширину 128 бит, что соответствует четырехсловной структуре по четыре адреса в каждом ряду. Память может быть сконфигурирована по желанию пользователя без специальной сегментации на память программ и память данных. Для доступа к данным процессор может адресовать одно, два или четыре слова и передавать их в/из одно(го) или оба(их) вычислительных устройства(в) за один такт. Кроме внутренней памяти архитектура TigerSHARC позволяет адресовать 4 Гслов.

Объединение данных процессоров по любой из архитектур (кольцо, гиперкуб и т. п.) позволяет создавать полнофункциональные нейроускорители. Так, нейровычислительная сеть может быть выполнена в виде двумерного массива (в общем случае четыре ряда по высоте и  $n$  колонн) процессоров, подключенных к интерфейс-

ным платам и с помощью хост-интерфейса к процессору общего управления. Каждый процессор в сети соединен с соседними, расположенными сверху и снизу относительно него, посредством четырех из шести имеющихся SHARC-линков. Остальные линки процессоров используются для частичного приема данных, необходимых в процессе вычислений. Архитектура обеспечивает масштабируемую сетевую процессорную модель с общей средой вычислений для каждого узла сети. Эта сеть подсоединяется посредством стандартного интерфейса разделяемой памяти к хост-процессору, который выполняет роль контрольного пункта системы.

### DSP фирмы Texas Instruments

Сигнальные процессоры компании *Texas Instruments* [8] разделяются на два класса: процессоры для обработки чисел с фиксированной точкой и процессоры для обработки чисел с плавающей точкой. Первый класс представлен тремя семействами процессоров, базовыми моделями которых являются соответственно TMS320C10, TMS320C20, TMS320C50. Второй класс включает процессоры TMS320C30, TMS320C40, TMS320C80, которые поддерживают операции с плавающей точкой и представляют собой мультипроцессорную систему, выполненную на одном кристалле, а семейство TMS320Cxxxx включает процессоры как с фиксированной, так и с плавающей точкой.

Процессоры старших поколений одного семейства наследуют основные архитектурные особенности и совместимы "снизу вверх" по системе команд (чего нельзя сказать о процессорах, входящих в разные семейства). Процессоры компании *Texas Instruments* обладают высокоскоростными интерфейсными подсистемами, поэтому их предпочтительнее использовать для тех задач, в которых требуется выполнение интенсивного обмена с внешними устройствами (коммуникационные системы, различного рода контроллеры).

**Процессор TMS320C80** фирмы *Texas Instruments* с производительностью в 2 млрд. операций в секунду представляет собой комбинацию из пяти процессоров, реализованных по MIMD (multiple-instruction, multiple-data) архитектуре. На одном кристалле реализованы одновременно две технологии — DSP и RISC, расположены один управляющий RISC-процессор и четыре 32-разрядных цифровых сигнальных процессора усовершенствованной архитектуры с фиксированной точкой (ADSP0-ADSP-3), обладающие высокой степенью конвейеризации и повышенной (до 64 бит) длиной слова инструкций, а это, в свою очередь, позволяет описывать сразу несколько параллельно выполняемых команд. Каждый из процессоров работает независимо друг от друга и может программироваться отдельно друг от друга и выполнять различные или одинаковые задачи, обмениваясь данными через общую внутрикристалльную кэш-память.

Суммарная производительность TMS320C80 на регистровых операциях составляет 2 млрд. RISC-подобных команд в секунду. Благодаря столь высокой производительности TMS320C80 может заменить при реализации приложений более 10 высокопроизводительных ЦСП или ЦП общего назначения. Пропускная способность шины TMS320C80 достигает 2,4 Гбайт/с в потоке данных и 1,8 Гбайт/с в потоке инструкций.

TMS320C80 обеспечивает высокую степень гибкости и адаптивности системы, построенной на его базе, которая достигается за счет наличия на кристалле параллельно функционирующих DSP-процессоров и главного RISC-процессора. Архитектура процессора TMS320C80 относится к классу MIMD (Multiple-Instruction, Multiple-Data) — множественный поток команд, множественный поток данных. Входящие в состав TMS320C80 процессоры программируются независимо один от другого и могут выполнять как разные, так и одну общую задачу. Обмен данными между процессорами осуществляется через общую внутрикристалльную память. Доступ к разделяемой внутрикристалльной памяти обеспечивает матричный коммутатор (Crossbar), выполняющий также функции монитора при обращении к одному сегменту памяти нескольких процессоров.

Большинство известных на сегодня нейровычислителей на базе DSP строятся на основе микропроцессоров семейства TMS320C4x. Благодаря своей уникальной структуре эти DSP получили широкое распространение в мультипроцессорных системах и практически вытеснили ранее господствующее в этой области семейство транспьютеров, производимых рядом европейских компаний. Процессоры TMS320C4x совместимы по системе команд с TMS320C3x, однако обладают большей производительностью и лучшими коммуникационными возможностями.

Все больше завоевывающее популярность, в том числе и для нейроприложений, семейство процессоров TMS320C6xxx обладает рекордной производительностью 1600 MIPS. Благодаря этому возможен принципиально новый взгляд на существующие системы связи и телекоммуникаций. Высокая производительность микропроцессоров обеспечивается благодаря новой архитектуре *VelocityTI™* с очень длинным командным словом (VLIW, Very Long Instruction Word). Архитектура *VelocityTI* образована множеством параллельно работающих процессоров, которые позволяют выполнять несколько инструкций за один командный цикл. Именно такой параллелизм архитектуры процессора обеспечивает высокую производительность.

### DSP фирмы Motorola

Сигнальные процессоры компании *Motorola* на сегодня в меньшей степени, чем рассмотренные выше, используются для реализации нейропарадигм. Они подразделяются на семейства 16- и 24-разрядных микропроцессоров с фиксированной точкой — DSP560xx, DSP561xx, DSP563xx, DSP566xx, DSP568xx, и микропроцессоры с плавающей точкой — DSP960xx. Линия 24-разрядных микропроцессоров компании *Motorola*, включает два семейства: DSP560xx и DSP563xx. Основные принципы, положенные в основу архитектуры сигнальных микропроцессоров *Motorola*, были разработаны и воплощены в семействе DSP560xx. Дальнейшие работы по совершенствованию сигнальных процессоров проводились по трем направлениям:

- наращивание производительности 24-разрядных процессоров за счет конвейеризации функциональных модулей и повышения тактовой частоты;
- создание дешевых 16-разрядных микропроцессоров с расширенными средствами взаимодействия с периферией;

## Характеристики нейрочипов

Наименование	Фирма-изготовитель	Разрядность, бит	Максимальное число синапсов*	Максимальное число слоев**	Примечание
MA16	Siemens	48 (умножители и сумматоры)	—	—	400 ММАС
NNP (Neural Networks Processor)	Accurate Automation	Nx16	—	—	MIMD, N — число процессоров
CNAPS-1064	Adaptive Solutions	16	128 Кбайт	64	—
100 NAP Chip	HNC	32	512 Кбайт	4	4 процессорных элемента
Neuro Matrix NM6403, Такт. частота 50 МГц	Модуль, Россия	64 (векторный процессор), 32 RISC ядро	4096 шт.	24	Совместим с портами TMS320C4x
Neuro Matrix NM6404, Такт. частота 133 МГц	Модуль, Россия	64 (векторный процессор), 32 RISC ядро	4096 шт.	~48	Совместим с портами TMS320C4x
CLNN 32	Bellcore	32	496	32 нейрона	10 <sup>8</sup> перекл./с
CLNN 64		64	1024		2 × 10 <sup>8</sup> перекл./с
NC 3001	NeuriGam	16	4096 шт.	32	—
ZISC 036 (Zero Instruction Set Computer)	IBM	64-разрядный входной вектор	—	36 нейронов	Частота 20 МГц, векторно-прототипный нейрочип
ETANN 80170NW	Intel	64 входа	Два банка весов 64 × 80	64 нейрона в слое, 3 слоя	Аналоговая
MD-1220	Micro Devices	16	64 шт.	8	8 нейронов
MT 19003 — Neural Instruction Set Processor	Micro Circuit Engineering (MCE)	16-разрядный умножитель, 32-разрядный сумматор	—	1	RISC МП с 7 специальными командами
Neuro Fuzzu	National Semiconductor	—	—	—	—
NI 1000	Nestor	5—16 (одного нейрона)	—	1024 прототипных 250-мерных векторов	Векторно-прототипный нейрочип
NLX420 (NLX 110, 230)	Adaptive Logic	16	1 Мбайт	16	16 процессорных элементов
OBL Chip	Oxford Computer	16	16 Мбайт	—	—
L-Neuro 1.0	Philips	16	—	16 нейронов	26 МГц
L-Neuro 2.3		16	1536	192 (12 × 16)	60 МГц
RSC (Speech Recognition Chip) — 164	Sensory Circuits	—	—	—	—
ORC 110xx (Object Recognizer Chip)	Synaptics	—	—	—	—
Pram-256 Chip	UCLi Ltd.	8 (одного нейрона)	—	256 нейронов	33 МГц
SAND	Datafactory	16	—	4	200 МСРС
ACC		16	—	—	—
Геркулес	Россия	16	1 Мбайт	64	—
Neuro Classifier	Университет Твена, DESY	70 вх. нейронов	—	6 (внутр.) 1 вх., 1 вых.	2 × 10 <sup>10</sup> перекл./с
ANNA	AT&T	Число нейронов 16—256	4096 весов	—	Число входов у нейрона 256—16
WSC (Wafer Scale Integration)	Hitachi	—	64 связи на нейрон	576 нейронов	—
SASLM2	Mitsubishi	2 (одного нейрона)	—	4096 (64 × 64) нейронов	50 МГц
TOTEM	Kent (Univer UK), di Trento (Италия)	16 (одного нейрона)	—	64 нейрона	30 МГц
Neuron 3120, Neurom 3150	Echelon (США)	8 бит (шина данных)	—	—	Наличие параллельных, последовательных и коммуникационных портов

Примечания: \* определяет размер внутрикристалльной памяти весов; \*\* определяется числом операций умножения с накоплением, выполняемых за один такт для операндов длиной 8 бит.

- ВСМП на основе принципиально новых сверхпараллельных нейросетевых алгоритмов решения различных задач (на базе нейроматематики).

Ориентация в выполнении нейросетевых операций обуславливает, с одной стороны, повышение скоростей обмена между памятью и параллельными арифметическими устройствами, а с другой стороны — уменьшение времени весового суммирования (умножения и накопления) за счет применения фиксированного набора команд типа регистр-регистр. Рассмотрим некоторые нейрочипы подробнее.

### Нейросигнальный процессор NeuroMatrix ("Модуль" [9], Россия)

Основой NeuroMatrix NM6403 (рис. 9, см. вторую сторону обложки) является процессорное ядро NeuroMatrixCore (NMC), которое представляет собой синтезобильную модель высокопроизводительного DSP-процессора с архитектурой VLIM/SIMD (язык Verilog). Ядро состоит из двух базовых блоков: 32-битного RISC-процессора и 64-битного векторного процессора, обеспечивающего выполнение векторных операций над данными переменной разрядности (патент РФ № 2131145). Имеются два идентичных программируемых интерфейса для работы с внешней памятью различного типа и два коммуникационных порта, аппаратно совместимых с портами ЦПС TMS320C4x, для построения многопроцессорных систем.

Базовыми для нейропроцессора являются вычисления вида:  $Z_i = f(Y_i) = f(U_i + \sum(X_j W_{ij}))$ , ( $i = 1, \dots, M; j = 1, \dots, N$ ), где  $Z_i$  — выходной сигнал  $i$ -го нейрона;  $X_j$  —  $j$ -й входной сигнал слоя;  $U_i$  — смещение  $i$ -го нейрона;  $W_{ij}$  — весовой коэффициент  $j$ -го входа  $i$ -го нейрона;  $Y_i$  — сумма взвешенных входов  $i$ -го нейрона;  $f$  — функция активации;  $N$  — число входных сигналов слоя;  $M$  — число нейронов в слое. Операнды  $Z_i$ ,  $X_j$ ,  $U_i$  и  $W_{ij}$  представлены в дополнительном параллельном коде и могут иметь произвольную разрядность.

Основными особенностями данного нейропроцессора являются:

- возможность работы с входными сигналами (синапсами) и весами переменной разрядности (от 1 до 64 бит), задаваемой программно, что обеспечивает уникальную способность нейропроцессора увеличивать производительность с уменьшением разрядности операндов;
- быстрая подкачка новых весов на фоне вычислений;
- 24 операции умножения с накоплением за один такт при длине операндов 8 бит;
- V-аппаратная поддержка эмуляции нейросетей большой размерности;
- реализация функции активации в виде пороговой функции или функции ограничения;
- наличие двух широких шин (по 64 разряда) для работы с внешней памятью любого типа: до 4 Мбайт SRAM и до 16 Гбайт DRAM;

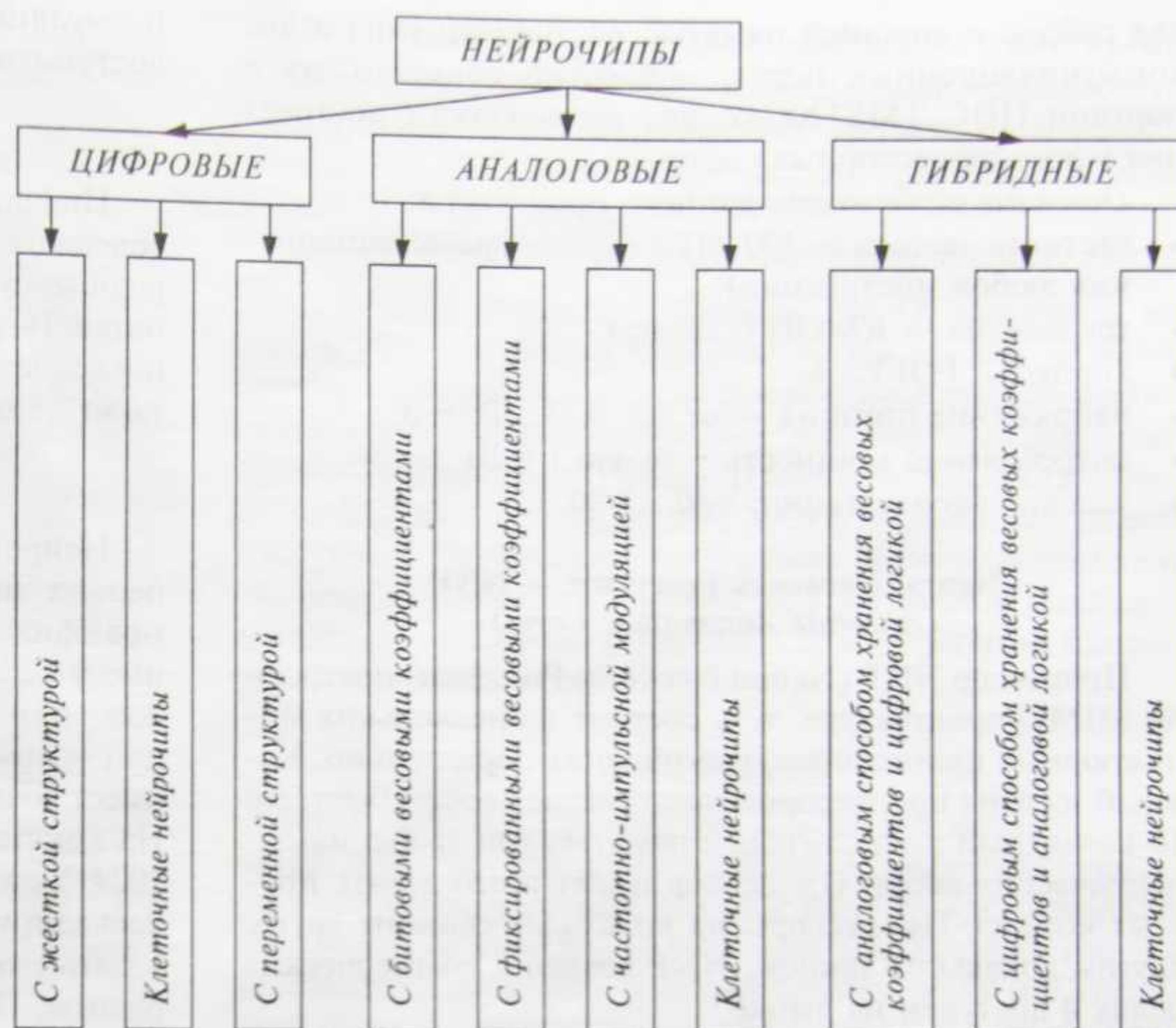


Рис. 8. Обобщенная классификация нейрочипов

- наличие двух байтовых коммуникационных портов ввода/вывода, аппаратно совместимых с коммуникационными портами TMS320C4x для реализации параллельных распределенных вычислительных систем большой производительности;
  - возможность работать с данными переменной разрядности по различным алгоритмам, реализуемым с помощью хранящихся во внешнем ОЗУ программ.
- Технические характеристики:*
- число вентилях на кристалле — 100 000;
  - размер кристалла — 10 × 10,5 мм при технологии 0,7 мкм;
  - потребляемая мощность — не более 3 Вт;
  - пиковая производительность для байтных операндов — 720 MCPS (миллионов соединений или умножений с накоплением в секунду) при тактовой частоте 30 МГц; при бинарных операциях — 8640 MCPS.

Нейропроцессор благодаря своей универсальности сможет применяться как базовый элемент для плат нейроускорителей PC, для создания нейрокомпьютерных параллельных вычислительных систем большой производительности, а также для аппаратной поддержки операций над матрицами большой размерности и в задачах цифровой обработки сигналов. Используется в нейроускорителях фирмы "Модуль" (Россия) [9].

### Процессор NeuroMatrix® NM6404

NeuroMatrix® NM6404 [9] представляет собой высокопроизводительный DSP-ориентированный RISC-микропроцессор. В его состав входят два основных блока: 32-разрядное RISC-ядро и 64-разрядный VECTOR-сопроцессор для поддержки операций над векторами с элементами переменной разрядности. NM6404 по системе команд совместим с предыдущей версией NM6403. Имеются два идентичных программируемых интерфейса

для работы с внешней памятью различного типа и два коммуникационных порта, аппаратно совместимых с портами ЦПС TMS320C4x, для возможности построения многопроцессорных систем.

Основные особенности данного процессора:

- тактовая частота — 133 МГц (8нс — время выполнения любой инструкции);
- технология — КМОП 0,25 мкм;
- корпус — PQFP256;
- напряжение питания — от 2,5 В, 3,3 В, 5 В;
- потребляемая мощность — около 1,0 Вт;
- условия эксплуатации: -40... +80 °С.

#### Neural Networks Processor — NNP (Accurate Automation Corp.)

Процессор NNP (Neural Networks Processor) построен по MIMD-архитектуре, т. е. состоит из нескольких миниатюрных процессоров, работающих параллельно. Каждый из этих процессоров представляет собой быстрый 16-разрядный вычислитель с памятью для хранения синаптических весов. Процессор имеет всего девять простых команд. Процессоры на кристалле связаны друг с другом локальной шиной. NNP создан в коммерческих целях и доступен на рынке.

В комплект поставки процессора включены средства разработки программ, а также библиотека подпрограмм с реализованными нейросетевыми алгоритмами, такими как сети Хопфилда, сети Кохенена и др. Процессор поставляется на платах под шины ISA, VME. Производительность — 140MCPS для однопроцессорной системы, 1,4GCPS для 10-процессорной системы.

#### Нейропроцессор MA16 (Siemens)

MA16 (рис. 10, см. вторую сторону обложки) изготовлен по технологии КМОП (1 мкм), состоит из 610 тыс. транзисторов и выполняет до 400 млн. операций умножения и сложения в секунду. Используется в качестве элементной базы нейрокомпьютера Synaps 1 и нейроускорителей Synaps 2 и Synaps 3 (распространяемых сегодня на рынке французской фирмой *Tiga Technologies*).

MA16 представляет собой программируемый каскадируемый процессор для векторных и матричных операций. Он поддерживает на аппаратном уровне следующие операции:

- матричное умножение;
- матричное сложение/вычитание;
- нормировка результата;
- вычисление векторной нормы (метрики L1 и L2);
- вычисление векторного расстояния (мера Манхэтэна, геометрическое расстояние).

Процессор содержит (рис. 11, см. вторую сторону обложки) четыре идентичных процессорных элемента, работающих параллельно. Входные данные имеют точность 16 бит, тактовая частота 50 МГц. Для операций матричного умножения/сложения скорость вычислений достигает  $8 \times 10^8$  операций/с. Программное обеспечение работает в среде UNIX/XWIND и реализовано на C++. Нейронная сеть тоже описывается на C++ или может вводиться интерактивно с помощью графического интерфейса типа OSF/Motif, что позволяет визуализировать конфигурацию чипа после отображения на него структуры сети. Хорошо развиты средства тестирования

и эмуляции. С 1995 года MA16 является коммерчески доступным продуктом.

#### MD1220 (Micro Devices)

Цифровой нейрочип MD1220 фирмы *Micro Devices* содержит восемь нейронов с восемью связями и 16-разрядные сумматоры. Во внутрикристалльной памяти хранятся 16-разрядные веса. Входы имеют одноразрядные последовательные умножители с продолжительностью такта 7,2 мкс. Средняя производительность около 9 MCPS.

#### L-Neuro (Philips)

Нейропроцессор L-Neuro фирмы *Philips* — один из первых нейропроцессоров. Широко известны две его модификации: L-Neuro 1.0 и L-Neuro 2.3. Вторая версия имеет 12 слоев, а первая — один слой из 16 одноразрядных, или двух 8-разрядных, или четырех 4-разрядных, или восьми 2-разрядных процессорных элементов, т. е. имеет возможность работать в мультиразрядном режиме. На кристалле реализован 1 Кбайт памяти для хранения 1024 8-разрядных или 512 16-разрядных весов. Гибкая каскадируемая структура нейрочипа позволяет использовать его при реализации различных нейросетевых парадигм. При реализации 64 8-разрядных процессорных элементов средняя производительность составляет 26 MCPS (32 MCUPS).

#### NLX-420 (NeuroLogix)

Каждый из 16 процессорных элементов нейрочипа NLX-420 фирмы *NeuroLogix* содержит 32-разрядный сумматор, логику параллельного выполнения 16 умножений. Средняя производительность 300 MCPS. Также имеется возможность каскадирования и мультиразрядных вычислений.

#### СБИС ETANN 80170NX (INTEL)

Аналоговая СБИС ETANN 80170NX фирмы *INTEL* содержит 64 входа, 16 внутренних уровней и 64 нейрона (пороговый усилитель с сигмоидной передаточной функцией). Каждый вход соединен с 64 синапсами. Передаточная функция нейрона в СБИС близка к сигмоиде.

Усиление передаточной функции определяет чувствительность нейрона. Низкое значение усиления позволяет интерпретировать выход нейрона как аналоговый, а высокое — как цифровой. Максимальное значение выхода нейрона определяется напряжением  $V_{tfo}$ . Веса ограничены интервалом  $[-2,5, 2,5]$ . Скорость прохождения сигнала по одному слою зависит от усиления и примерно равна 1,5 мкс, что и определяет быстродействие. Точность выполнения операций примерно эквивалентна 6 бит, быстродействие — 1,3—109 переключений/с. Обучение выполняется методом Back Propagation с помощью Intel Neural Network Training System (INNTS). Применяемое системное окружение представляет собой специальную версию пакета DynaMind. Обучение выполняется до получения приемлемого уровня ошибки выхода сети, и после достижения удовлетворительной работы веса загружаются в СБИС. Для реальной работы такого обучения недостаточно, так как программа симуляции не может точно смоделировать аналоговую работу СБИС и, например, не отслеживает флуктуации в передаточной функции каждого нейрона. Поэтому следующий этап

обучения представляет собой так называемый chip-in-loop (CIL) training, когда после каждого цикла веса записываются в СБИС и выход сети непосредственно используется в процессе обучения. Точность ETANN — 5—6 разрядов для весов и выходов.

Поскольку ETANN представляет собой аналоговую СБИС, то для ее надежной работы важны стабильные внешние условия. Специально сконструированный для этого модуль обеспечивает низкую пульсацию источника питания < 5 мВ (напряжение питания 5 В) и температурную стабильность.

#### СБИС CLNN32/CLNN64 (Bellcore)

Гибридный нейрочип CLNN32 состоит из 32 нейронов и 496 двунаправленных адаптивных синапсов. CLNN64 содержит только 1024 адаптивных синапсов. В наборе CLNN32/CLNN64 все нейроны взаимосвязаны, так что любая топология сети отображается подбором синапсов. Динамика сети полностью аналоговая, но значения синапсов хранятся/обновляются в цифровом виде с точностью 5 бит. На аппаратном уровне реализовано обучение сети — подбор весов происходит по алгоритму обучения машины Больцмана или Mean Field. Внутри также имеется некоррелированный генератор шума (32 канала), используемый при обучении по методу машины Больцмана. CLNN32 может быть использован независимо или совместно с CLNN64 для построения более сложной архитектуры сети. Производительность достигает  $10^8$  переключений/с (при работе с CLNN64 удваивается). Для CLNN32 это означает, что примерно  $10^5$  32-бит образцов/с или 32 аналоговых канала (с полосой пропускания 50 кГц) могут быть использованы для быстрого распознавания/обучения. Время распространения для одного слоя нейронов < 1 мкс. "Охлаждение" (по методу Больцмана) или MF-обучение требует 10—20 мкс. CLNN32 по сравнению с СБИС ETANN имеет следующие очевидные преимущества:

- быстрое обучение (микросекунды по сравнению с часами при CIL-процессе);
- эффективный алгоритм обучения Больцмана, обеспечивающий быстрое нахождение "почти оптимального" решения;
- простые и быстрые процедуры чтения/записи весов, выполняемые в цифровом виде, что значительно увеличивает скорость обмена между сетевым сервером и клиентами в сети;
- легкая каскадируемость.

#### СБИС ANNA (AT&T)

Другим примером реализации гибридного нейрочипа является нейрочип ANNA. Логика нейрочипа — цифровая, хранение весов — аналоговое (на элементах динамической (конденсаторной) памяти). Чип содержит 4096 весов, максимальное число нейронов 256. Точность весов — 6 разрядов, для однослойной сети  $64 \times 64$  производительность достигает 2.1 GCPS.

#### СБИС NeuroClassifier

Аналоговая СБИС NeuroClassifier создана в университете Твенте совместно с DESY. Ее архитектура состоит из входного слоя (70 входов, полоса пропускания до 4 Гбайт/с), шести внутренних слоев и одного выходного

нейрона. Точность аналогового умножения 5 бит, время решения всего лишь 20 нс, что позволяет использовать NeuroClassifier в триггере первого уровня. Эквивалентное быстродействие примерно  $2 \times 10^{10}$  переключений/с.

#### SAND/1 (Simple Applicable Neural Device)

Компания Datafactory (бывшая INCO) выпустила на рынок SAND/1 (Simple Applicable Neural Device). SAND/1 представляет собой каскадно соединенные систолические процессоры, оптимизированные для быстрого решения задач в нейросетевом базисе. Производительность одного процессора составляет 200 MCPS (миллионов связей в секунду). Процессор имеет четыре 16-битных потока и 40-битный сумматор. SAND/1 был разработан Исследовательским центром в Карлсруе и Институтом микроэлектроники Штутгарта.

#### Inova N64000

Этот нейрочип фирмы Inova SIMD-архитектуры относится так же, как и предыдущий, к классу систолических нейропроцессоров. Он содержит 80 процессорных элементов, из которых 64 образуют основную матрицу, а 16 являются резервом, 4 Кбайт памяти весов и 32 регистра общего назначения. Арифметический модуль процессора имеет девять параллельных 16-разрядных умножителей и один 32-разрядный сумматор.

#### 100 NAP (Hecht-Nielson Computer)

Другой систолический нейрочип 100 NAP фирмы Hecht-Nielson Computer содержит четыре 32-разрядных процессорных элемента с плавающей точкой. Средняя производительность около 150 MFLOPS, адресуемое адресное пространство внекристальной памяти — 512 Кбайт.

#### MT19003 (Micro Circuit Engineering)

Нейрочип MT19003 также относится к классу систолических нейропроцессоров. Основой архитектуры является RISC-ядро с семью специальными командами, 16-разрядный векторный умножитель и 32-разрядный сумматор, внутрикристальная память для хранения весов отсутствует. Точность входов и весов — 13 разрядов. Средняя производительность — 50 MCPS.

#### NEURON (Echelon, США)

Нейропроцессор NEURON ориентирован на создание кластерно-параллельных вычислительных систем. Программно-алгоритмическое обеспечение по управлению кластерной структурой реализовано внутри кристалла. Предложенная архитектура кристалла стала в настоящее время основой стандарта ANSI/EIA 709.1—1999 построения различных АСУ технологическими процессами [2].

В семействе нейрочипов NEURON выделяют NEURON 3120 и NEURON 3150. Структурная схема NEURON 3150 приведена на рис. 12. Кристалл содержит 2К динамической памяти для хранения весов и данных, 512 байт (EEPROM) для размещения управляющих программ. Для выполнения специализированных сетевых и управляющих операций в структуре кристалла имеется два спецвычислителя: Applications CPU, Network CPU. Также следует отметить широкие коммуникационные возможности, реализованные на кристалле.

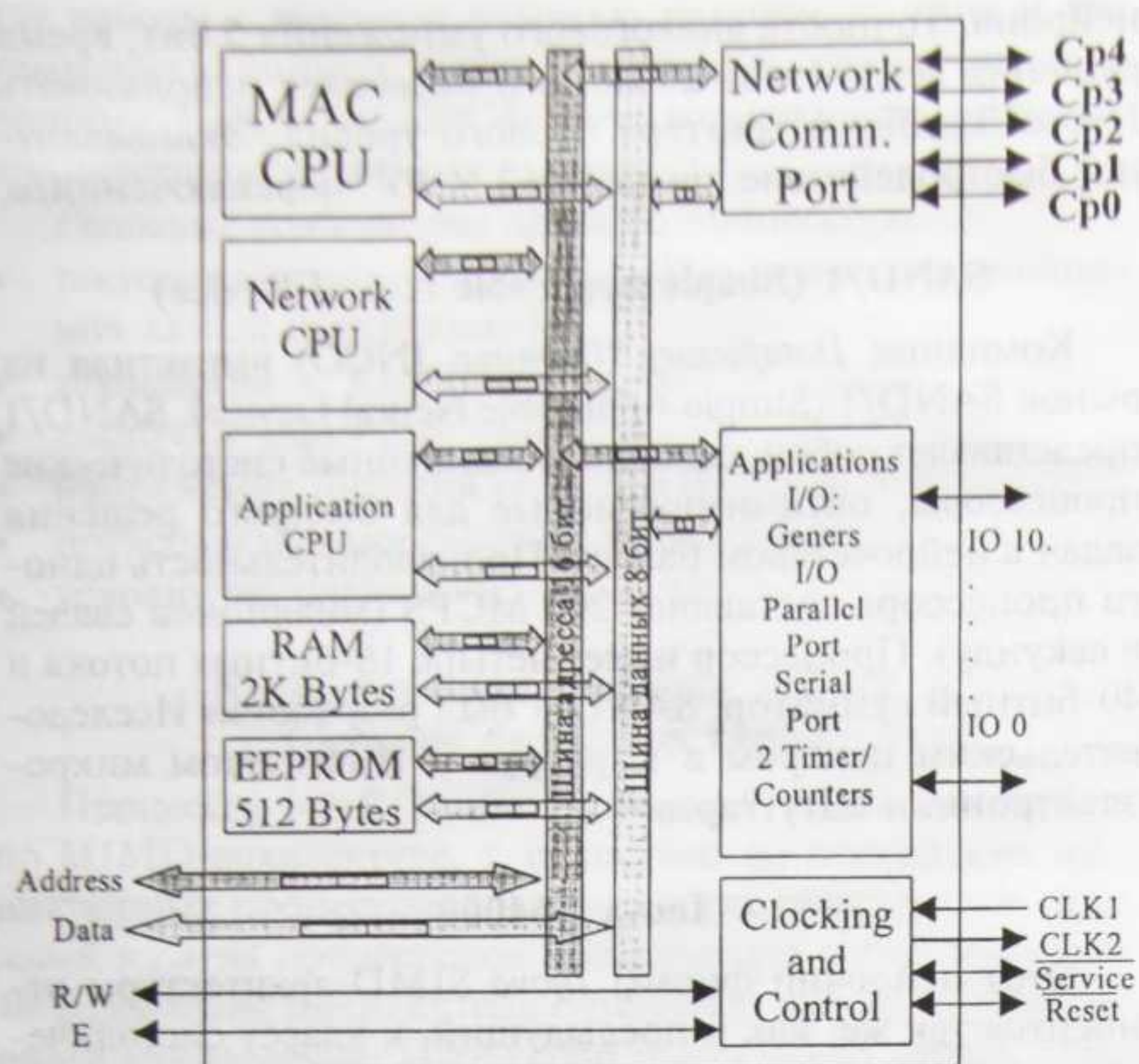


Рис. 12. Структурная схема нейрочипа NEURON 3150 фирмы Echelon (США)

### ZISC036 (IBM)

Нейрочип ZISC036 (Zero Instructions Set Computer) относится к нейрочипам векторно-прототипной архитектуры, т. е. алгоритм обучения строится на соотношении входного вектора и весов входов нейронов, запомненных прототипными векторами. Он содержит 36 нейронов. Ориентирован на решение широкого круга задач, в том числе задач распознавания образов и классификации. Каждый нейрон представляет собой независимый процессор. ZISC способен решать и real-time-задачи (рис. 13).

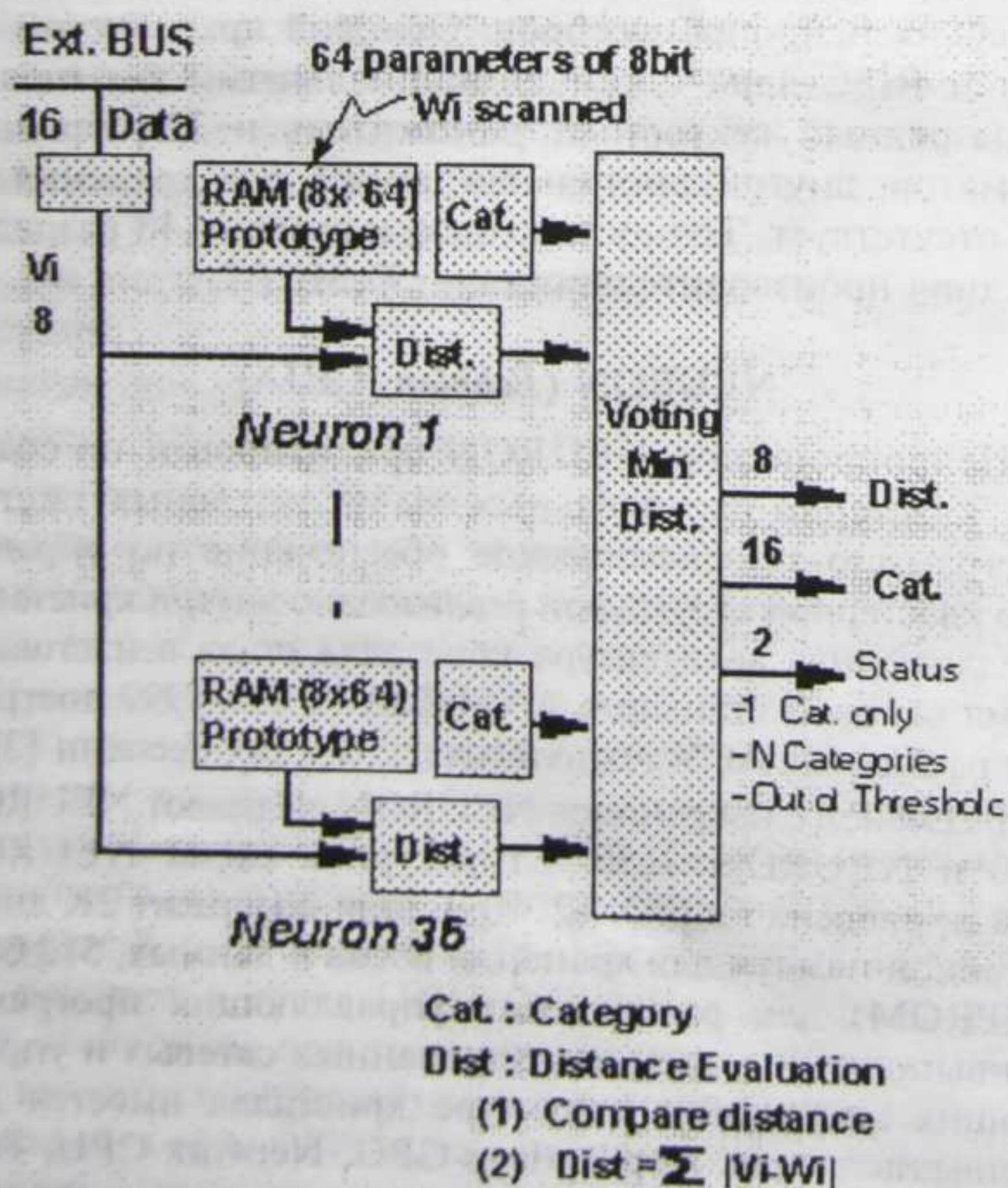


Рис. 13. Функциональная схема процессора ZISC

### Характеристики ZISC036:

- 36 нейронов;
- возможность соединения нескольких процессоров (каскадирования);
- от 1 до 64 компонентов во входном векторе;
- напряжение питания 5 В;
- потребляемая мощность 1 Вт при 16 МГц;
- частота 0—20 МГц;
- CMOS-технология;
- входной вектор загружается через 3,5 мкс, результат появляется через 0,5 мкс.

Для увеличения производительности фирма IBM разработала ISA- и PCI-модули параллельно работающих ZISC-процессоров.

В заключение краткого обзора нейрочипов приведем сводную таблицу (табл. 4).

Таблица 4

### Производительность нейропроцессов (средняя)

Наименование нейрочипа	Конфигурация	CPS	CPSPW	CPPS	CUPS
NLX420	32—16,8 bit mode	10M	20K	640M	—
100 NAP	4 chips, 2M wts, разрядность мантиссы 16 бит	250M	125	256G	64M
WSI (Hitachi)	576 neuron Hopfield	138M	3.7	10G	—
N64000 (Inova)	64—64—1,8 бит	871M	128K	56G	220M
MA16 1 ZISC036	chip, 25 МГц 64 8 бит (разрядность входного вектора)	400M	15M	103G	—
MT19003	4—4—1, 32 МГц	32M	32M	6.8G	—
MD1220	8-8	9M	1M	142M	—
NI 1000	256 5 бит (разрядность входного вектора)	40 000 vec. in sec.	—	—	—
L-neuro-1	1-chip, 8 бит	26M	26K	1.6G	32M
NM6403	8 бит, 50 МГц	1200M	150M	77G	—

## ЧАСТЬ 3. Аппаратная реализация нейровычислителей

Рассмотрев в предыдущих разделах основы нейроматематики и элементную базу нейровычислителей, остановимся более подробно на анализе структурно-функционального построения нейровычислителей.

### 3.1. Нейроускорители на базе ПЛИС

Построение нейровычислителей на базе ПЛИС, с одной стороны, позволяет гибко реализовать различные нейросетевые парадигмы, а с другой, сопряжено с большими проблемами разводки всех необходимых межсоединений. Выпускаемые в настоящее время ПЛИС (с числом вентилях от 5 до 100 тысяч) имеют различные функциональные возможности. Нейровычислители на базе ПЛИС, как правило, позиционируются как гибкие нейровычислительные системы для научно-исследовательских целей и мелкосерийного производства. Для построения более производительных и эффективных ней-

ровычислителей требуется применение сигнальных процессоров.

Вопросам создания нейровычислителей на ПЛИС посвящено большое число работ, представленных на прошедшей выставке "Нейрокомпьютеры и их применение". Мы в качестве примера рассмотрим нейровычислитель, созданный в НИИ Системных исследований (РАН) [16].

**Параллельный перепрограммируемый вычислитель (ППВ)** (рис. 14, см. вторую сторону обложки) разработан в стандарте VME и реализован на базе перепрограммируемых микросхем семейства 10K фирмы *Altera*. Вычислитель предназначен для работы в качестве аппаратного ускорителя и является ведомым устройством на шине VME. Он должен включаться в систему как подчиненное устройство основной управляющей ЭВМ (host-машины) с универсальным процессором. Тактовая частота вычислителя 33 МГц [16].

Основные характеристики представлены в табл. 5.

Таблица 5

Основные характеристики ППВ

Название алгоритма	Pentium-100, с	Pentium-II-333, с	Ultra SPARC, с	ППВ, с
Свертка с ядром 4 × 4	0.65	0.11	0.76	0.02
Медианный фильтр	1.97	0.49	0.75	0.001
Повышение контрастности	0.51	0.13	1.31	0.004
Прямое поточечное сравнение с маской 32 × 32	43.78	7.14	58.89	0.142
Поиск локальных неоднородностей 32 × 32	0.120	0.028	0.146	0.032
Умножение матрицы на матрицу	8.61	0.60	12.31	0.011

Примечание. Оценки приведены для: Pentium-100 при частоте 100 МГц, объем ОЗУ 16 Мбайт; Pentium-333 при частоте 350 МГц, объем ОЗУ 128 Мбайт; UltraSPARC при частоте 200 МГц, объем ОЗУ 64 Мбайт. Тактовая частота ППВ 33 МГц.

ППВ используется для построения систем распознавания образов на основе обработки телевизионной, тепловизионной и другой информации, а также систем, основанных на реализации алгоритмов с пороговыми функциями и простейшими арифметическими операциями, и позволяет добиться значительной скорости вычислений. Вычислитель состоит из следующих функциональных блоков [16]:

- схемы управления;
- базовых вычислительных элементов (БВЭ);
- контроллера внешней шины (контроллер E-bus);
- контроллера системной шины (контроллер VME);
- двух массивов статической памяти (ОЗУ0, ОЗУ1);
- блока высокоскоростных приемников/передатчиков.

Схема управления используется для управления БВЭ и потоками данных в вычислителе и представляет собой простейший RISC-процессор. Структура и набор команд процессора могут изменяться в зависимости от типа решаемой задачи.

С точки зрения программиста, вычислитель можно представить как RISC-процессор (схема управления или управляющий процессор) и шесть векторных процессоров

(вычислительных элементов), обрабатывающих SIMD-команды (одна команда для многих данных).

Методика быстрого создания нейровычислителей на ПЛИС приведена в [17]. Основные тенденции в проектировании нейровычислителей на ПЛИС — это увеличение плотности компоновки нейрокристаллов за счет уменьшения площади межсоединений и функциональных узлов цифровых нейронов. Для решения этой задачи находят применение:

- оптические связи для передачи информации между нейронами;
- модификаций программно-аппаратной реализации функциональных элементов для нейровычислений;
- оптимизация представления промежуточных данных в слоях нейронов — нейросети со сжатой формой внутренних данных.

### 3.2. Нейроускорители на базе каскадного соединения сигнальных процессоров

Такие нейровычислители представляют собой мультипроцессорные системы с возможностью параллельной обработки, что позволяет реализовывать на их основе нейровычислительные системы, в структуре которых можно выделить две основные части:

- управляющую Host-ЭВМ, реализованную на основе обычной вычислительной системы с CISC- или RISC-микропроцессорами;
- виртуальное (или внешнее) аппаратное средство, подключаемое к Host-ЭВМ посредством внутренних (внешних) системных интерфейсов, выполняющее основные вычислительные операции.

Остановимся на особенностях аппаратной реализации таких нейровычислителей (НВ).

В основе построения НВ данного типа лежит использование сигнальных процессоров, объединенных между собой в соответствии с определенной архитектурой, которая обеспечивает параллельность выполнения вычислительных операций. Как правило, такие НВ строятся на основе гибкой модульной архитектуры, которая обеспечивает простоту конфигурации системы и наращиваемость вычислительной мощности путем увеличения числа процессорных модулей или применения более производительных сигнальных процессоров (рис. 15). НВ данного типа реализуются главным образом на базе несущих модулей стандартов ISA, PCI, VME. Основными их функциональными элементами являются модуль матричных процессоров (МСП), рабочая память, память программ, модуль обеспечения ввода/вывода аналоговых и цифровых сигналов (АЦП, ЦАП, ТТЛ) и разъем для подключения БДП, БИКП, БКИ.

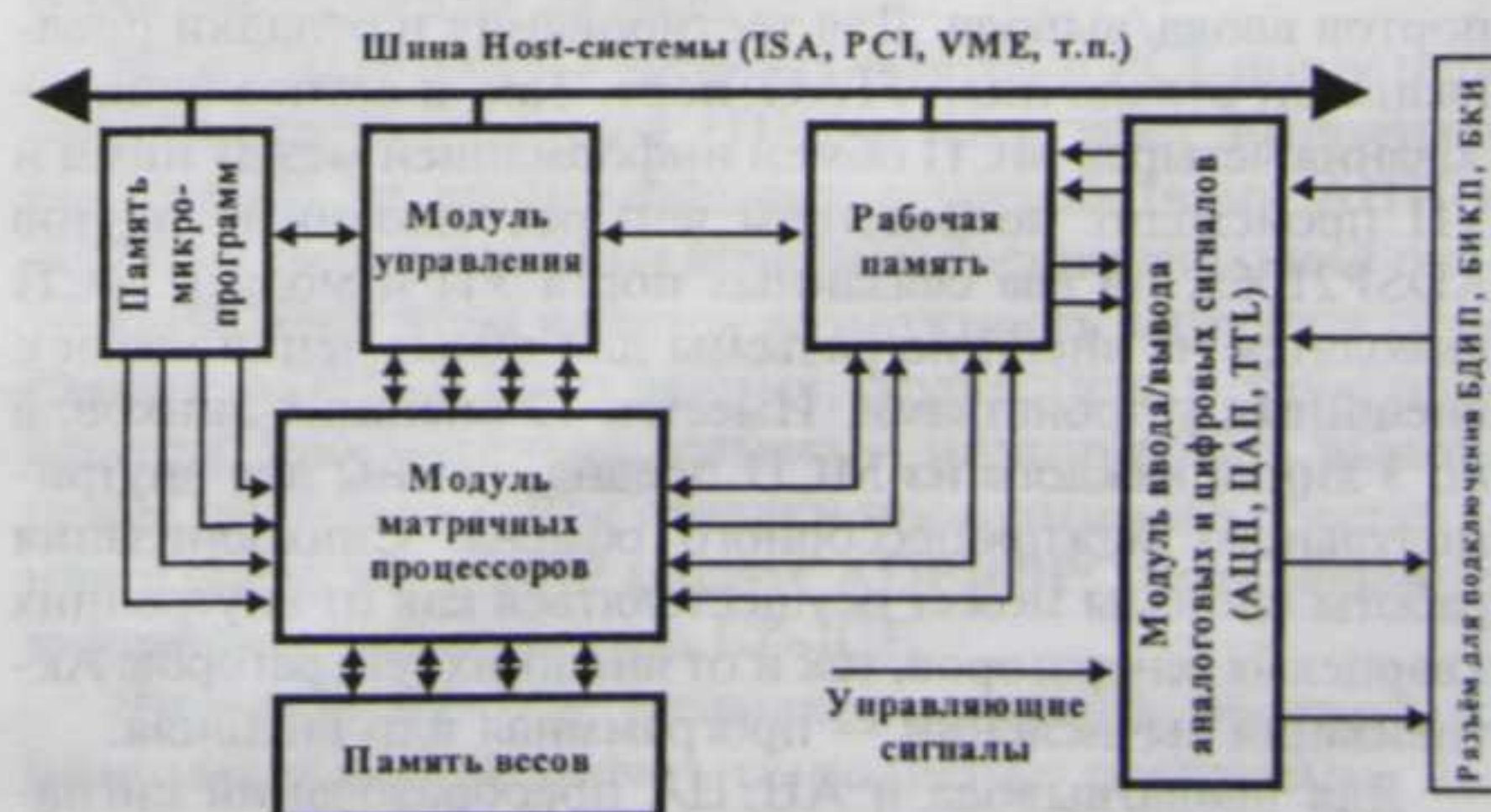


Рис. 15. Обобщенная функциональная схема виртуального НВ



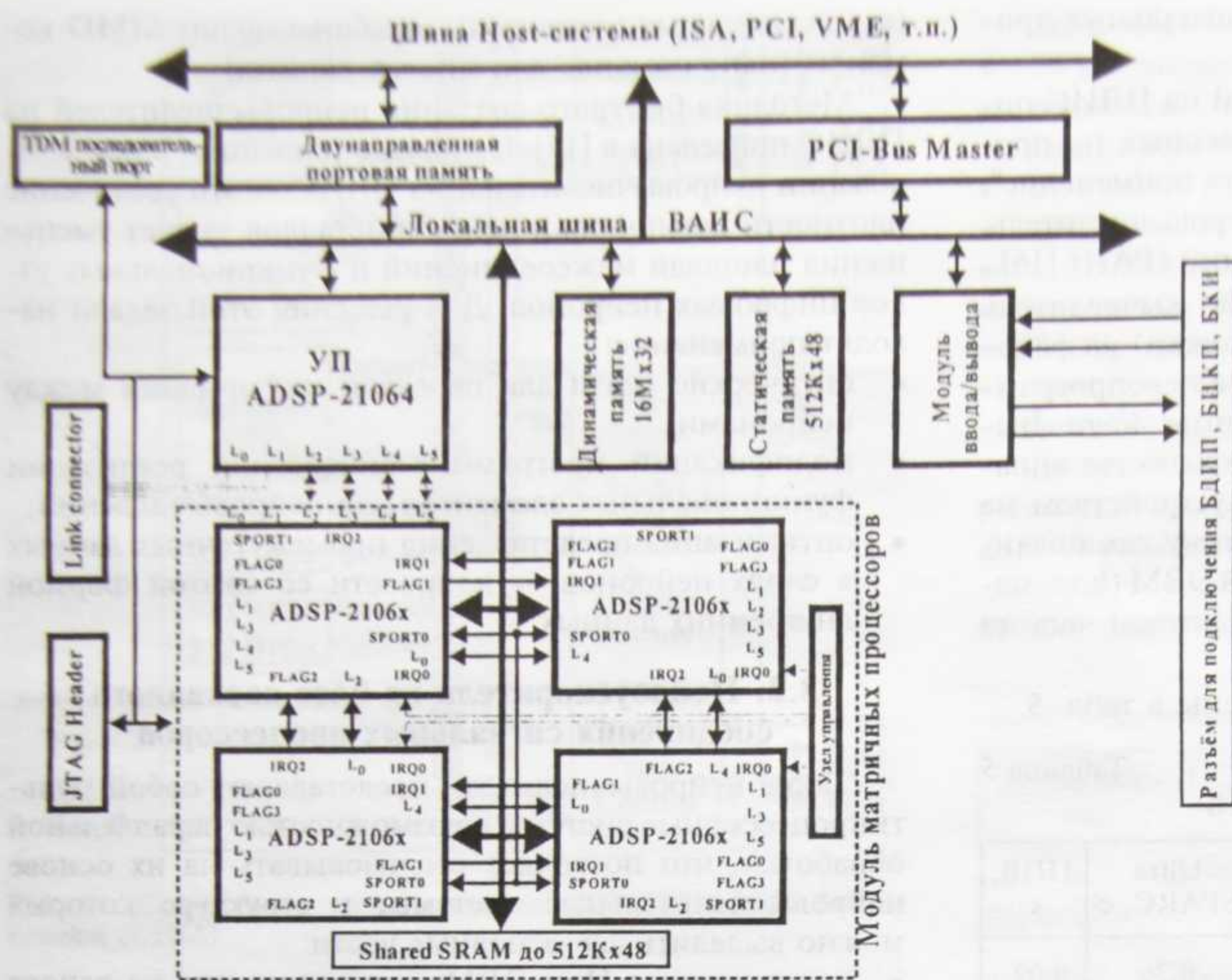


Рис. 16. Реализация НВ на основе ADSP2106x

да/вывода сигналов (включающий АЦП, ЦАП и TTL линии), а также модуль управления, который может быть реализован на базе специализированного управляющего сигнального процессора (УП) и ПЛИС или иметь распределенную структуру, при которой функции общего управления распределены между МСП.

Для построения НВ данного типа наиболее перспективным является использование сигнальных процессоров с плавающей точкой ADSP2106x, TMS320C4x, 8x, DSP96002 и др.

Типовая структурная схема реализации НВ на основе сигнальных процессоров ADSP2106x приведена на рис. 16 [5]. В ее состав включены один управляющий сигнальный процессор для осуществления функций общего управления и до восьми процессоров, осуществляющих параллельные вычисления согласно заложенным алгоритмам (матричные сигнальные процессоры).

Управляющий и матричные процессоры образуют кластер процессоров с общей шиной и ресурсами разделяемой памяти. Обмен информацией между управляющим процессором, матричными процессорами, Host-ЭВМ и внешней средой осуществляется посредством портов ввода/вывода. Для тестирования и отладки предназначен отладочный JTAG-порт. Так, в случае использования четырех МСП обмен информацией между ними и УП происходит посредством четырех связанных портов ADSP2106x, по два связанных порта УП и модуля МСП выводятся на внешние разъемы для обеспечения связи с внешними устройствами. Имеется 12 внешних линков, а по 3 линка каждого из МСП предназначены для внутримодульного межпроцессорного обмена. Синхронизация работы системы может осуществляться как от внутренних кварцевых генераторов, так и от внешних генераторов. Активизация вычислений — программная или внешняя.

Для ввода/вывода и АЦ/ЦА преобразований сигналов предназначен специализированный модуль, кото-

рый включает в себя: универсальный цифровой TTL-порт, АЦП, ЦАП, узел программируемых напряжений для смещения шкал АЦП и установки порога срабатывания стартовых компараторов, узел фильтрации выходных аналоговых сигналов, подсистему тестирования, узел синхронизации и управления, буферную память FIFO. Первоначальная загрузка осуществляется по Host-интерфейсу или по линкам. Управляющий интерфейс любого МСП позволяет управлять процессорным сбросом и прерываниями, его идентификационным номером и т. п.

Такая архитектура НВ обеспечивает выполнение операций цифровой обработки сигналов (ЦОС) в реальном времени, ускорение векторных вычислений, возможность реализации нейросетевых алгоритмов с высоким параллелизмом выполнения векторных и матричных операций.

Несколько DSP, входящих в структуру НВ образуют распределенную вычислительную структуру из процессорных модулей, соединенных между со-

бой высокоскоростными портами. Данный вариант реализации НВ может быть построен с использованием от двух до восьми сигнальных процессоров.

При использовании двух параллельных 32-разрядных DSP TMS320C40 обмен информацией при реализации нейросетевых алгоритмов осуществляется с помощью шести связанных портов с пропускной способностью в 30 Мбайт/с и каналов DMA каждого из процессоров. Поддерживая параллельную независимую работу, подсистема DMA и процессор обеспечивают параллельный обмен информацией со скоростями до 560 Мбайт/с. С помощью высокоскоростных портов возможна реализация на основе данных DSP таких архитектур, как кольца, иерархические деревья, гиперкуб и т. п. Каждая из локальных шин TMS320C40 обеспечивает обмен информацией на скоростях до 120 Мбайт/с.

Процессорные модули функционируют независимо и при необходимости объединяются посредством связанных портов. Функции обмена, управления процессорными модулями, прерываниями и каналами DMA реализуют ПЛИС, например фирмы Xilinx. Применение в НВ динамических реконфигурируемых структур (нейросети со структурной адаптацией) и использование последних ПЛИС семейств xC2xxx—xC4xxx (фирмы Xilinx), или аналогичных, требует минимизации времени на реконфигурацию ПЛИС, которые чаще всего программируются в режимах Master Serial и Peripheral. Основным недостатком при использовании данных режимов перепрограммирования заключается в зависимости процесса переконфигурации ПЛИС от встроенного тактового генератора. Минимальные потери времени можно получить при проведении переконфигурации ПЛИС в режиме Slave Serial, в котором внутренний тактовый генератор отключен, а синхронизация осуществляется посредством внешних синхросигналов. Реконфигуратор ПЛИС выполняется в виде специализированной микросхемы (например, XC2018-84pin-50MHz, XC3020-68pin-50MHz).

Рассмотренные варианты НВ обеспечивают выполнение ЦОС и нейроалгоритмов в реальном масштабе времени, ускорение векторных и матричных вычислений по сравнению с традиционными вычислительными средствами в несколько раз и позволяют реализовывать нейросеть с числом синапсов до нескольких миллионов.

Еще больше повысить производительность НВ данного типа можно при использовании одного из самых мощных на сегодня сигнальных процессоров — TMS320C80, TMS320C6xxx фирмы *Texas Instruments*.

Примером реализации нейровычислителя на DSP фирмы *Motorolla* является нейровычислитель NEURO TURBO фирмы *Fujitsu*. Он реализован на основе четырех связанных с кольцом 24-разрядных DSP с плавающей точкой MB86220 (основные параметры: внутренняя точность — 30 разрядов, машинный цикл — 150 нс, память программы — 25К слов × 2 (внутренняя), 64К слов × 4 (внешняя), технология изготовления — КМОП 1,2 мкм). Активационная функция нейронов ограничивается в диапазоне от 0 до 1, а возможные значения входов не превышают 16 разрядов, что обуславливает достаточную точность при 24-разрядной архитектуре. Построение нейрокомпьютера на основе кольцевой структуры объединения DSP позволяет снизить аппаратные затраты на реализацию подсистемы централизованного арбитража межпроцессорного взаимодействия.

Нейрокомпьютер NEURO TURBO состоит из четырех DSP, связанных друг с другом посредством двухпортовой памяти (ДПП). Каждый из DSP может обращаться к двум модулям такой памяти (емкостью 2К слов каждая) и к рабочей памяти (РП) (емкостью 64К слов × 4 Банка) в своем адресном пространстве. Вследствие того, что доступ ДПП осуществляется случайным образом одним из соседних DSP, передача данных между ними происходит в асинхронном режиме. Рабочая память используется для хранения весовых коэффициентов, данных и вспомогательной информации. Для успешной работы НС необходимо получение сверток во всех элементарных нейронных узлах. Кольцевая структура объединения DSP обеспечивает конвейерную архитектуру свертки, причем передача данных по конвейеру осуществляется посредством ДПП. После того, как DSP загружает данные из одной ДПП, он записывает результаты своей работы в смежную ДПП, следовательно, кольцевая архитектура параллельной обработки обеспечивает высокую скорость операции с использованием относительно простых аппаратных решений.

Для выполнения функций общего управления используется Host-ЭВМ на основе обычной вычислительной системы. Обмен данными между нейроплатой и Host-ЭВМ происходит через центральный модуль ДПП. Загрузка программ в DSP осуществляется посредством памяти команд для каждого DSP. Следовательно, его архитектура полностью соответствует параллельной распределенной архитектуре типа MIMD. Пиковая производительность системы 24 MFLOPS.

Для реализации модели НС иерархического типа фирмой *Fujitsu* выпущена нейроплата на основе DSP MB86232 с собственной памятью до 4 Мбайт, что позволяет осуществлять моделирование НС, содержащей более 1000 нейронов. Структура НС включает в себя входной, промежуточный и выходной уровни [наибольшее число скрытых слоев — два (ограничение по памяти)].

Для обучения нейрокомпьютера используются оригинальные фирменные алгоритмы: алгоритм виртуального импеданса, алгоритм скорректированного обучения и алгоритм расширенного обучения.

Каждая из рассмотренных типовых структур реализации НС может быть промоделирована на основе рассмотренных выше вариантов построения мультипроцессорных НВ. Так, для НВ на основе TMS320C4x при реализации какой-либо из рассмотренных архитектур (кольцо, иерархическое дерево, гиперкуб и т. п.) достаточно только изменить назначения коммуникационных портов, что обеспечивает гибкость и масштабируемость при исследовании и разработках нейросетевых систем различной архитектуры.

#### **Высокопараллельные вычислители и нейроускорители производства АОЗТ "Инструментальные системы" (Россия)**

АОЗТ "Инструментальные системы" [11] выпускает в России самую широкую номенклатуру различных многофункциональных мультипроцессорных плат как в виде "виртуальных" устройств, так и в виде крейтовых модулей и полнофункциональных вычислителей. Элементной базой высокопараллельных нейроускорителей данной фирмы являются DSP фирмы *Analog Devices* (SISD Sharc, SIMD Sharc, Static Superscalar Sharc) и DSP фирмы *Texas Instruments* (в основном семейств TMS320C4x, TMS320C6xxx). Интерфейсы выпускаемых модулей различны: ISA, PCI, CompactPCI, VME, IEEE 1394, RS232, RS482, RS485 и др. Спектр заказных изделий практически неограничен. Проанализируем линейку серийно выпускаемых высокопараллельных ускорителей фирмы "Инструментальные системы" на конкретных примерах.

**Модуль ADP160PCI.** Эта плата является наиболее ярким представителем линейки параллельных виртуальных вычислителей, основанная на процессоре ADSP-21160 и предназначенная для построения систем радиосвязи, радиолокации, гидролокации, вибродиагностики, обработки изображений. Она построена на базе 32-разрядного 600 Mflops процессора SHARC2 компании *Analog Devices*. Сигнальный процессор ADSP-21160 работает на тактовой частоте 100 МГц, имеет два исполнительных устройства для выполнения SIMD-операций и внутреннюю память 4 Мбит. 64-разрядная шина процессора осуществляет одновременную выборку двух 32-разрядных данных для их параллельной обработки в процессоре. Процессор обеспечивает вычисление 1024-точечного комплексного преобразования Фурье с битреверсными перестановками за 90 мкс.

Плата ADP160PCI устанавливается в PCI-шину персонального компьютера (ПК) и позволяет выполнять программы во взаимодействии с устройствами АЦП и ЦАП, размещенными на дочерней расширительной плате ADM. При этом имеется возможность загружать программы ADSP-21160 и данные, осуществлять сброс процессора, просматривать память и инициировать выполнение программы. Для отладки программного обеспечения в среде VisualDSP к плате ADP160PC подключается внутрисхемный эмулятор EZ-ICE.

Эмулятор EZ-ICE позволяет загружать программы, начинать и останавливать выполнение программы, наблюдать и изменять состояние регистров и памяти, осуществлять другие операции отладки.

Блок-схема платы ADP160PCI представлена на рис. 17 (см. третью сторону обложки). Обращения к процессору ADSP-21160 осуществляются по интерфейсу шины PCI. Так как интерфейс PCI непосредственно связан с Host-портом процессора, ПК получает доступ к его внутренним ресурсам. Пропускная способность моста при записи из ПК во внутреннюю память процессора ADSP-21160 составляет 15 Мбайт/с, при чтении ПК из внутренней памяти процессора ADSP-21160 составляет 10 Мбайт/с. Чтобы разгрузить процессоры от рутинной работы по пересылке массивов, в микросхему моста встроены два контроллера прямого доступа к памяти, обеспечивающие пересылку данных между памятью ПК и внутренней памятью процессора ADSP-21160 со скоростью до 40 Мбайт/с.

**Модуль DSP60V6.** Это высокопроизводительный мультипроцессорный модуль сбора и цифровой обработки сигналов, основанный на процессоре цифровой обработки сигналов ADSP 21060/62 SHARC. Он позволяет выполнять программы SHARC во взаимодействии с устройствами, размещенными на дочерней плате ADM, в качестве которой могут использоваться модули АЦП, ЦАП. ADP60V5 устанавливается в промышленные крейты с размером плат 6U.

Модуль работает как автономно, так и с компьютером, имеющим шину VME. Можно загружать программы ADSP-21060/62 и данные через шину VME и/или через пользовательские выходы разъема J2/P2 (X2), осуществлять сброс процессоров, просматривать память и инициировать выполнение программ.

Модуль имеет в своем составе процессорный кластер из шести процессоров ADSP2106x производительностью 120 MFLOPs каждый. В процессорном кластере устанавливается до 1М × 48 бит оперативной статической памяти и до 16М × 32 бит оперативной динамической памяти. Кластер имеет в своем адресном пространстве VME-интерфейс и FLASH-память 4М × 8 бит. Модуль может работать независимо от шины VME, в этом случае прием и передача данных проводится по шести коммуникационным портам.

**Модуль ADP44PCI.** Несущий процессорный модуль ADP44PCI ориентирован на цифровую обработку сигналов и может быть легко конфигурирован для решения конкретных задач путем загрузки в модуль прикладного программного обеспечения и, в случае необходимости, подключения высокоскоростного submodule аналогового ввода/вывода и/или модуля стандарта TIM-40. Несколько ADP44PCI образуют распределенную вычислительную структуру из процессорных модулей, соединенных между собой с помощью высокоскоростных связанных портов. В состав вычислительной структуры входят высокоскоростные устройства аналогового ввода/вывода, процессоры, локальная и глобальная память, интерфейсы ПЭВМ.

Процессорный несущий модуль ADP44PCI использует параллельный 32-разрядный DSP с плавающей точкой TMS320C44 фирмы *Texas Instruments*. Четыре 30 Мбайт/с связанных порта и канала ПДП процессора обеспечивают разнообразные возможности высокоскоростного обмена. Две шины TMS320C44 обеспечивают обмен 120 Мбайт/с каждая. Работая совместно, но независимо, процессор и ПДП TMS320C44 обеспечивают параллельный обмен данными со скоростью до 560 Мбайт/с.

Высокоскоростные связанные порты позволяют объединять процессоры в кольца, деревья, гиперкуб и др.

Процессорный несущий модуль ADP44PCI удовлетворяет стандарту TIM-40 и выполнен с использованием программируемых логических интегральных схем фирм *Xilinx* и *Altera*, реализующих функции обмена, средства управления процессором, прерываниями и каналами ПДП. Процессор TMS320C44 имеет производительность 60 MFlops. Объем статической памяти составляет от 128К × 32 до 512К × 32 на двух шинах процессора, предусмотрено расширение путем установки модулей SIMM динамической памяти до 16М × 32 на глобальной шине процессора. Собственно кристалл TMS320C44 имеет 8 Кбайт встроенной памяти. На глобальной шине размещен интерфейс ADM-Connect для связи с submodule аналогового ввода/вывода со скоростью до 100 Мбайт/с.

Промышленный стандарт TIM-40 разработан консорциумом под руководством фирмы *Texas Instruments* для модулей на C40 и C44. Спецификация определяет аппаратные ресурсы, способы программирования и организации многопроцессорных систем. Процессорный несущий модуль ADP44PCI удовлетворяет основным требованиям стандарта TIM-40, что позволяет организовывать многопроцессорные системы с другими модулями стандарта TIM-40.

Важными составляющими модуля стандарта TIM-40 являются: каналы связи с ведущей ПЭВМ, межпроцессорного обмена по коммуникационным портам; описание доступных C40 ресурсов в специальном идентификационном ПЗУ IDROM, порядок загрузки исполняемых программ, габариты и порядок подсоединения TIM-40 модулей.

Внутренняя структура интерфейса ПЭВМ модуля ADP44PCI полностью определяется загруженной конфигурацией ПЛИС. Реализация иной внутренней структуры системы для решения специальных задач пользователя возможна по отдельному заказу.

**Модули на базе TMS320C6701.** Модули данного семейства ориентированы на применение в телекоммуникационных системах, включая базовые станции, системах медицинской диагностики, навигационных системах и системах позиционирования, мультимедиа и т. п., где требуется сверхвысокая вычислительная мощность. Структурная схема мультипроцессорной платы на базе TMS320C6701 приведена на рис. 18.

**Модуль ADP6701PCI** фирмы "Инструментальные системы" (рис. 19, см. третью сторону обложки) построен на базе микропроцессора TMS320C6701, обладающего производительностью 1 GFLOPS. Имеет в своем составе восемь параллельных вычислительных блоков, обеспечивает выполнение БПФ на 1024 отсчета за 109 мкс.

Отличительными особенностями платы является: буферная память тракта ввода 64К × 32, вывода 64К × 16, двухпортовая статическая память 64К × 32, синхронная динамическая память до 16 Мбайт. Благодаря гибкому аппаратному интерфейсу на ПЛИС обеспечивается программная совместность различных submodule.

### Нейроускорители НТЦ "Модуль"

За период 1989—2000 гг. в НТЦ "Модуль" разработан ряд образцов нейрокомпьютерных вычислительных систем на современной элементной базе ведущих зарубежных фирм и на основе нейропроцессора *NeuroMatrix*

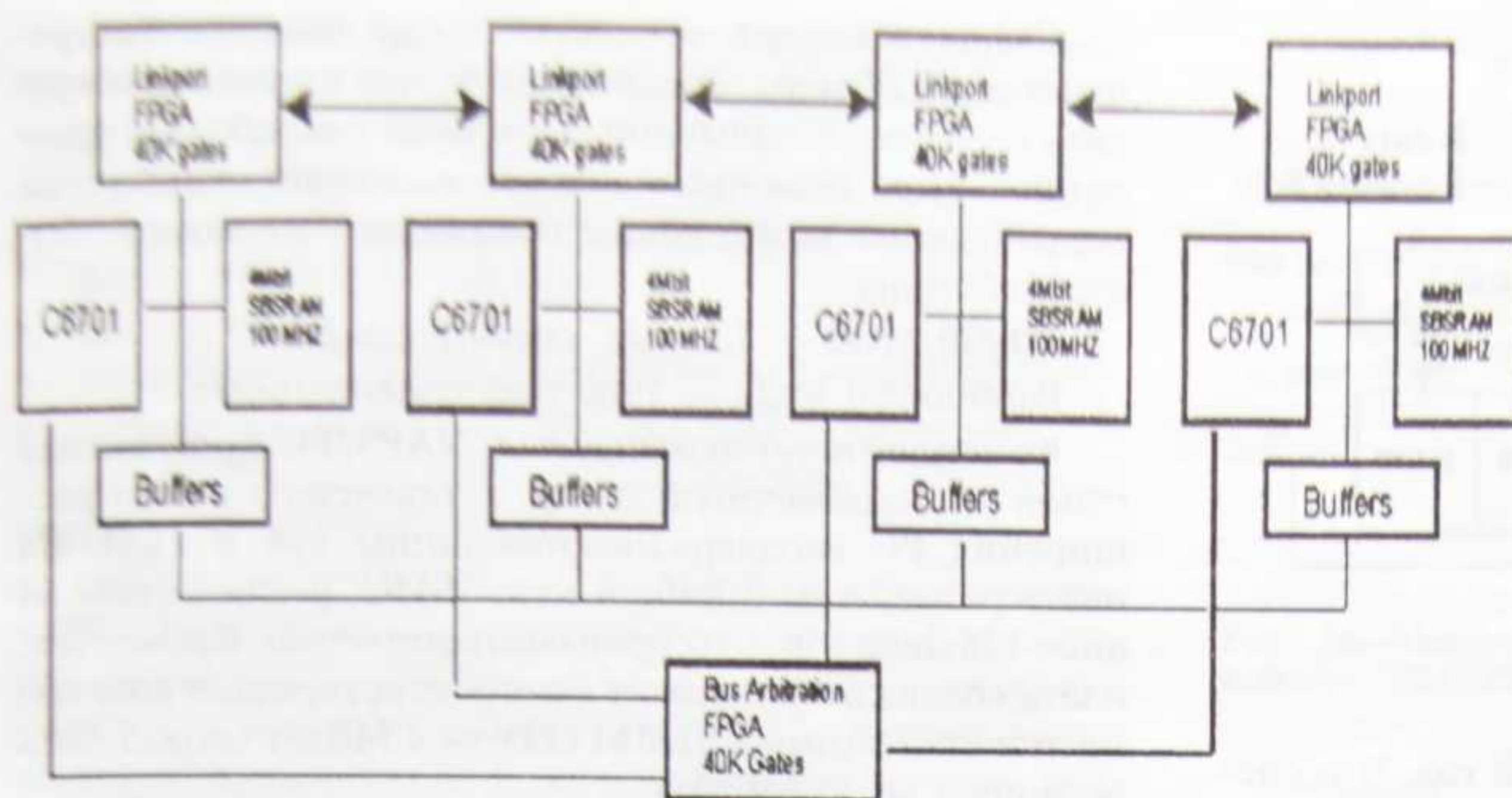


Рис. 18. Структурная схема мультипроцессорного вычислителя на базе TMS320C6701

6403 и 6404, созданного в НТЦ "Модуль" [10]. Среди широкой линейки различных ускорителей можно выделить: нейроускорители на основе универсальных сигнальных процессоров (главным образом используется TMS320C4x) и нейроускорители на базе специализированного нейро-сигнального процессора NeuroMatrix, созданного в НТЦ "Модуль". Примерами ускорителей первого типа являются ускорительные платы: M1, M2 и т. п.

**Ускорительная плата M2.** Многопроцессорный модуль M2 цифровой обработки сигналов выполнен на цифровых сигнальных процессорах TMS320C40 фирмы *Texas Instruments* и представляет собой одноплатную многопроцессорную вычислительную машину. Модуль предназначен как для автономной работы, так и для функционирования в составе ПЭВМ с системной шиной VMEbus, в том числе состоящей из нескольких таких же модулей. Конструктивно блок выполнен в соответствии с механическим стандартом VMEbus IEEE 1014 (6 U).

Модуль M2 содержит:

- VMEbus контроллер;
- master/slave интерфейс;
- до шести TMS320C40 с частотой 50 МГц;
- до 2 Мбайт SRAM на процессор;
- до 64 Мбайт DRAM на плате;
- FLASH EEPROM до 0,5 Мбайт;
- внешние связи (скорость — 20 Мбайт/с);
- JTAG-интерфейс;
- RS-232-интерфейс.

Общая производительность — до 300 MFLOPS. Структурная схема ускорителя M2 приведена на рис. 20 (см. третью сторону обложки).

### 3.3. Нейровычислители на базе нейрочипов

Кроме нейроускорителей на базе ПЛИС и DSP в последнее время все большее распространение находят нейроускорители на базе специализированных нейросигнальных и нейросетевых процессоров. Проанализируем особенности их реализации на конкретных примерах.

#### Нейровычислители НТЦ "Модуль"

**Двухпроцессорный встраиваемый модуль МЦ4.01 (NM1).** Встраиваемый модуль МЦ4.01 (NM1) (рис. 21, см. третью сторону обложки) цифровой обработки сиг-

налов производства НТЦ "Модуль" [10] предназначен для решения различных задач нейронными и нейроподобными алгоритмами, а также задач цифровой обработки сигналов и ускорения векторно-матричных вычислений. Модуль выполнен на спроектированных и разработанных в НТЦ цифровых сигнальных процессорах NeuroMatrix® NM6403 и представляет собой одноплатный нейроускоритель. Конструктивно выполнен в виде платы, вставляемой в стандартный слот шины PCI.

Нейроускоритель содержит:

- два нейропроцессора NM6403;
- от 2 до 8 Мбайт статической памяти (SRAM);
- 64 Мбайта динамической памяти (EDO DRAM);

- четыре внешних коммуникационных порта с темпом обмена 20 Мбайт/с каждый.

Производительность:

- векторные операции — 1,9 млрд. операций с байтовыми операндами в секунду;
- скалярные операции — до 320 млн. операций в секунду;

Конструктивное выполнение:

- стандарт PCI (версия 2-1) с темпом обмена до 132 Мбайт/с.

**Нейроускоритель МЦ4.02.** Содержит один процессор NM6403 и обладает масштабируемой производительностью от 40 до 11 500 ММАС, обеспечивает обработку данных переменной разрядности от 1 до 64 бит. Модуль предназначен для работы в составе комплекса с системной шиной PCI, блок статической памяти модуля доступен для записи и чтения как со стороны процессора, так и со стороны шины PCI. На внешние разъемы модуля выведены два коммуникационных порта, аппаратно совместимых с портами TMS320C4x. Соединение коммуникационных портов нескольких модулей позволяет создавать мультипроцессорные системы различной конфигурации. Сейчас ведутся разработки по созданию модулей на базе процессора NM6404.

#### Нейровычислители Synapse

Компания *Siemens Nixdorf Informationssysteme* (SNI) — дочернее предприятие концерна *Siemens* в сотрудничестве с Маннгеймским университетом создала нейрокомпьютер под названием "Synapse 1", который появился на рынке в середине 1994 г., в дальнейшем были выпущены нейроускорители "Synapse 2" и "Synapse 3". Таким образом, SNI стала первой европейской фирмой, выпустившей нейрокомпьютеры. В настоящее время нейроускорители распространяются французской фирмой *Tiga Technologies* (gouimenou.tiga@wanadoo.fr).

Сфера применения нейромодулей — распознавание речи, изображений, образов, ускорение работы программных эмуляторов. Сложность моделирования на рабочей станции процесса самообучения для нейронных сетей до сих пор тормозила разработку нейронных применений, поскольку каждый шаг в обучении требует много времени. Что касается "Synapse", то за один час самообучения он достигает таких же результатов, что и нейрон-

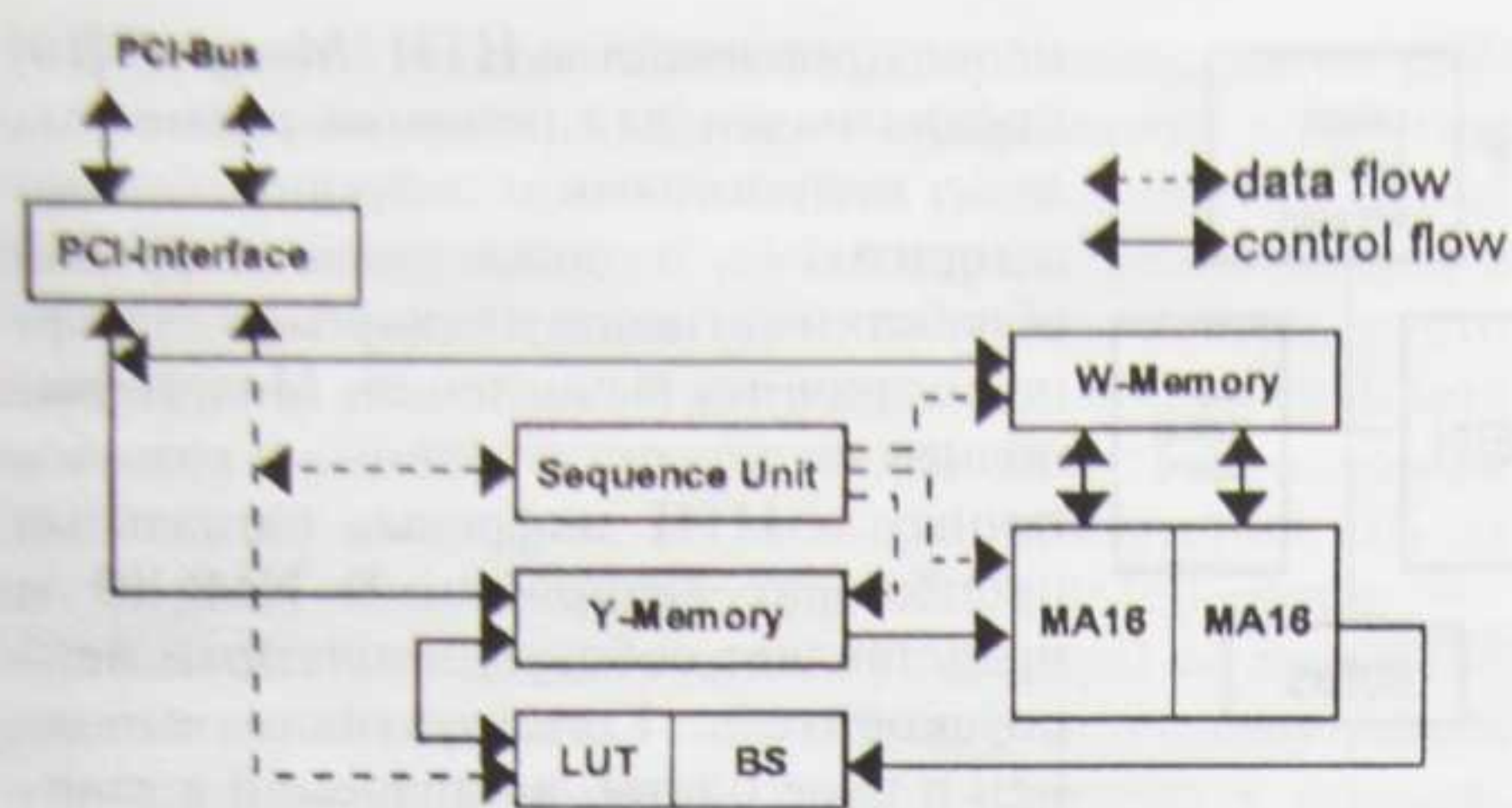


Рис. 23. Структурная схема "SYNAPSE-3"

ные сети в обычном компьютере за целый год. Эти системы обладают скалярной многопроцессорной архитектурой и наращиваемой памятью.

В состав "Synapse 2" (рис. 22, см. третью сторону обложки) входит один нейрочип MA16 (40 Гц), управляющий сигнальный процессор TMS320C50 (55 МГц), модуль целочисленной обработки на базе TMS320C50 (55 МГц), память образцов (Y-память), память весов (W-память).

Серийно выпускаемые нейроускорители "Synapse 3" поставляются с двумя процессорами MA16. Структурная схема нейроускорителя "Synapse 3" представлена на рис. 23. Пиковая эффективность одной нейроплаты "Synapse 3" — 2,4 млрд. операций/с. Программное обеспечение работает в среде UNIX/XWIND и реализовано на C++. Нейронная сеть тоже описывается на C++ или может вводиться интерактивно с помощью графического интерфейса типа OSF/Motif, что позволяет визуализировать конфигурацию СБИС после отображения на нее структуры сети. Средства тестирования и эмуляции развиты очень хорошо.

Сравнительная диаграмма производительности нейроплат и Pentium200 на матричных операциях приведена выше на рис. 24.

#### Нейроплата CNAPS/PC-128 фирмы Adaptive Solutions

Плата CNAPS/PC была анонсирована в США в ноябре 1995 г. Она обладает рядом характеристик, которые на первый взгляд кажутся фантастическими. При обучении на сложных данных нейроплата превосходит ПК на базе Pentium по быстродействию в несколько сотен (на отдельных тестах — до 1000) раз. Нейроплата позволяет прогнозировать заведомо "гиблые" задачи — текущую ситуацию на мировом валютном рынке, динамику политических событий в регионах и даже исход футбольных матчей.



Рис. 24. Сравнительная диаграмма производительности матричных операций

Старшая модель семейства имеет пиковое быстродействие 2,27 млрд. соединений/с, что позволяет сократить время аналитической обработки данных. Для сравнения — при решении нейросетевых задач стандартные персональные компьютеры показывают следующее быстродействие:

486/50 MHz — 750 тыс. соединений/с.

Pentium/90 MHz — 1980 тыс. соединений/с.

Конструктивно нейроплата CNAPS/PC представляет собой полноразмерную плату, вставляемую в слот расширения PC (поддерживаются шины ISA и PCI). На плате размещены 2 либо 4 нейроБИС, реализующие 64 либо 128 нейропроцессоров соответственно. Кроме того, плата содержит 512 Кбайт быстродействующей кэш-памяти и стандартный SIMM ОЗУ — 4 Мбайт (может быть расширен до 36 Мбайт).

#### Нейроускорители фирмы IBM

**ISA-ускоритель с 16 x ZISC 036 (IBM).** ZISC/ISA предназначен для IBM PC-совместимых компьютеров. Ускоритель построен на 16 ZISC036 процессорах и имеет 576 нейронов. Возможна установка нескольких карт, одна из которых работает в режиме Master, а другие — Slave.

**PCI-ускоритель на ZISC 036.** IBM PCI ZISC-ускоритель — высокопроизводительный нейровычислитель, предназначенный для работы в PCI слоте. Характеристики ускорителя:

- рабочая частота . . . . . 33 МГц;
- производительность . . . . . 165000 операций/с;
- может содержать . . . . . 1, 7, 13 или 19 ZISC036.

#### 3.4. Нейрокомпьютеры

Рассматривая подходы к аппаратной реализации нейровычислителей, необходимо отметить, что несмотря на широкое распространение различных высокопараллельных ускорителей для различных задач число моделей полнофункциональных нейрокомпьютеров невелико, а коммерчески доступны из них единицы. Это и понятно, так как большинство из них реализованы для спецприменений. Наиболее яркими примерами нейрокомпьютеров являются [1, 2]: нейрокомпьютер "Synapse 1" (Siemens, Германия), нейрокомпьютер "Силиконовый мозг" (созданный в США по программе "Электронный мозг", предназначен для обработки аэрокосмических изображений, производительность 80 пэтафлоп ( $80 \times 10^{15}$  операций/с, объем равен объему мозга человека, потребляемая мощность 20 Вт), нейрокомпьютер "Эмбрион" (Россия) и нейрокомпьютеры, созданные в научном центре "Нейрокомпьютеры" (г. Москва).

#### Нейрокомпьютеры фирмы "Siemens"

**"SYNAPSE-1".** Базовый комплект "SYNAPSE-1" — N110 предполагает наличие главной ЭВМ — рабочей станции SUN SPARCSTATION 5 модели TX1 в качестве вспомогательного консолидирующего устройства, облегчающего процессы программирования, проектирования нейросетей, тестирования, управления внешними устройствами, вывода результатов и т. п. Главная ЭВМ сопрягается с аппаратурой "SYNAPSE-1" через шину VME.

В архитектуре "SYNAPSE-1" можно выделить четыре основных компонента: матричный процессор, память

весов, устройство управления и устройство данных. Основные характеристики нейрокомпьютера:

- процессорная плата с матрицей из восьми сигнальных процессоров MA16 с производительностью 3,2 млрд. операций умножения ( $16 \times 16$  бит) и сложения (48 бит) в секунду;
- память весов 128 Мбайт;
- устройство управления на базе MC68040 ("Motorolla");
- устройство данных на базе MC68040 ("Motorolla").

Все аппаратные средства размещаются в небольшом корпусе  $667 \times 398 \times 680$  мм.

Проведенные исследования показали, что по производительности выполнение нейросетевых операций на нейрокомпьютере "SYNAPSE-1" по крайней мере на три порядка превышает производительность традиционных вычислительных систем и позволяют моделировать нейросети с количеством синапсов, равным 64 миллионам, а гибкость архитектуры практически не ограничивает разнообразность реализуемых нейросетевых парадигм.

### Нейрокомпьютер "ЭМБРИОН"

Нейрокомпьютер "ЭМБРИОН" (рис. 25, см. третью сторону обложки) [12] разработан под руководством чл.-кор. МАИ, канд. техн. наук В. Д. Цыганкова ([sinteg@mail.ru](mailto:sinteg@mail.ru)). Известно несколько модификаций данного нейрокомпьютера для различных приложений: датчик случайных многомерных управляемых импульсных потоков "ЭМБРИОН-1", интересенсорный перенос "глаз—рука", техническая диагностика неисправностей энергогенератора самолетной электростанции ("ЭМБРИОН-2"), управление нестационарным объектом (ЛА) в реальном масштабе времени ("ЭМБРИОН-3" и "ЭМБРИОН-4"), орган технического зрения ("ЭМБРИОН-5"), управление тактильно осязательным адаптивным промышленным роботом "УНИВЕРСАЛ-5А" при обслуживании карусельной плавильной печи на стекольном заводе ("ПОИСК-1"), управление тактильно осязательным адаптивным промышленным роботом "Р-2" с искусственными мышцами при сборке и покраске ("ПОИСК-2"), управление тактильно осязательным мобильным автономным роботом "КРАБ-1" при взаимодействии с неориентированными предметами и др.

Виртуальный нейрокомпьютер "ЭМБРИОН" — модель мозга человека [12], возбуждение квазинейронов в нем создает виртуальное квантовое когерентное волновое поле, которое имеет прямое отношение к проблеме создания квантового нейрокомпьютера (КНК). Нейрокомпьютер "ЭМБРИОН" представляет собой некоторую разновидность квантовой макросистемы или квантово-механической системы.

На рис. 26 [12] представлена структурная схема процесса генерации нейронной сети в НК "ЭМБРИОН". Сигналы из внешней среды  $S$  проецируются на сенсорную матрицу  $SM$ , которая в модели НК представляет некоторое подобие проекционной зоны коры мозга. Под воздействием активирующего потока импульсов из блока выдвижения гипотез  $BVG$  — подкорковой структуры мозга, ретикулярной формации или некоторого внутреннего генератора активности процессов в нейронной сети, информация из  $SM$  переносится в регистр внутренней памяти  $PO$  и меняющиеся во времени его коды состояния формируют виртуальную квазинейронную сеть (НС) и ее выходную реакцию.

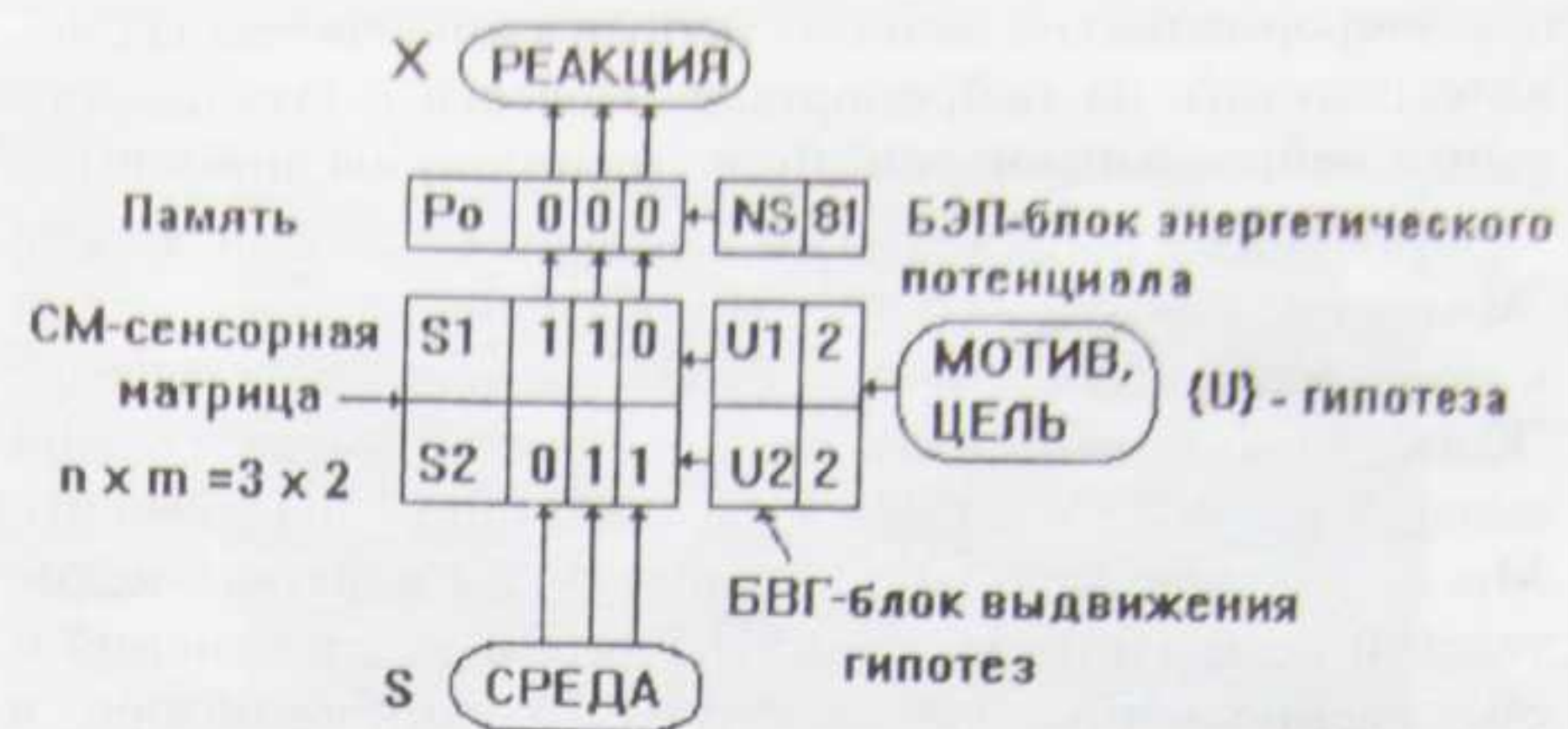


Рис. 26. Условия синтеза квазинейронной сети в нейрокомпьютере "ЭМБРИОН" [12]

Нейрокомпьютер "ЭМБРИОН" — это система, основанная на коллективном спиновом резонансе. Число частиц в нейрокомпьютерной квантовой системе точно известно и определяется разрядностью  $n$  НК. Блок выдвижения гипотез представляет собой устройство, формирующее и управляющее последовательностью действий унитарных преобразований пространства кубитов. Протокол квантовой криптографии в НК или алгоритм  $U$ -кратного применения унитарного оператора-выстрела  $|U| = 1$  — это процесс переноса информации из  $SI$ -й строки сенсорной матрицы в регистр внутренней памяти  $PO$ . Выполняется четвертое требование к КНК. Момент измерения результата  $Y$  — это момент времени  $t$ , когда  $U_k = U_{\max}$ . Выполняется последнее из вышеприведенного перечня пятое требование к КНК.

НК "ЭМБРИОН" представляет собой квантовый когерентный нейрокомпьютер и требует для своего воплощения совершенно другую, по сравнению с обычным ПК, функциональную элементную базу или "мозговую массу".

### Заключение

Нейроускорители и нейрокомпьютеры являются мощным средством проведения вычислений в реальном масштабе времени. Если раньше большая часть времени уходила на подготовку и проверку одной-единственной гипотезы, то теперь система обрабатывает данные и выдает заключения практически в реальном времени. Несмотря на свои впечатляющие возможности нейровычислители не очень распространены на рынке вследствие высокой цены (от единиц до десятков тысяч долларов для нейроускорителей и от десятков до сотен тысяч долларов для нейрокомпьютеров) и из-за специфики освоения. Основная причина отсутствия полнофункциональных нейровычислителей на рынке — закрытость разработок. Из сотен фирм, производящих специализированные нейрочипы, лишь единицы поставляют свою продукцию на массовый рынок. Остальные обслуживают военный комплекс либо создают единичные продукты для спецприложений.

К сожалению, из-за ограничений в объемах в рамках одного обзора нет возможности рассказать подробно обо всех известных на сегодня нейровычислителях и нейрокомпьютерах, в дальнейшем мы постараемся периодически возвращаться к данной теме и обсуждать вопросы вновь появившихся архитектурно-схематических решений и элементной базы нейровычислителей. Оператив-

ную информацию по данному направлению вы всегда сможете получить на нейропортале "Новости с Российского рынка нейрокомпьютеров" (<http://neurnews.iu4.bmstu.ru>).

Материалы данного обзора являются основой курсов "Микропроцессоры в системах управления" и "Архитектура и схемотехника современных ЭВМ", читаемых на кафедре "Конструирование и технология производства электронной аппаратуры" МГТУ им. Н. Э. Баумана (<http://iu4.bmstu.ru>). Мы будем признательны разработчикам нейровычислителей и всем заинтересованным лицам за дополнения и обсуждение вопросов, затронутых в данном обзоре, а также готовы к выполнению совместных проектов и внедрению ваших нейросетевых разработок в учебный процесс МГТУ им. Н. Э. Баумана. На сегодня основными направлениями деятельности кафедры в области нейросетевых приложений являются: создание нейроадаптивных систем активного управления пространственными волновыми полями (акустика, вибрации), нейроадаптивные системы управления робототехническими технологическими комплексами, распознавание изображений, обработка сигналов в системах неразрушающего контроля, создание систем интеллектуального управления трафиком в телекоммуникационных системах, исследовательские работы в области нейроинформатики (экспертные системы, аналитические системы, контекстно-поисковые системы и т. п.).

#### Список литературы

1. **Проблемы** построения и обучения нейронных сетей // Библиотечка журнала "Информационные технологии". 1999. № 1. 105 с.
2. **Галушкин А. И.** Некоторые исторические аспекты развития элементной базы вычислительных систем с массовым па-

раллелизмом (80-е и 90-е годы) // *Нейрокомпьютер*, 2000. № 1. С. 68—82.

3. **Горбань А. Н., Россиев Д. А.** Нейронные сети на персональном компьютере. Новосибирск: Наука, Сибирская издательская фирма РАН, 1996. 276 с.

4. **Кирсанов Э. Ю.** Цифровые нейрокомпьютеры: Архитектура и схемотехника / Под ред. А. И. Галушкина. Казань: Казанский государственный университет. 1995. 131 с.

5. **Власов А. И.** Аппаратная реализация нейровычислительных управляющих систем // *Приборы и системы управления*. 1999. № 2. С. 61—65.

6. <http://www.module.ru>

7. <http://www.analog.com>

8. <http://www.ti.com>

9. <http://www.ins-s.ru>

10. **Алюшин М. В.** Аппаратная реализация быстродействующих нейросетей на основе программируемой логики фирм AMD, ALTERA, XILINX // *Нейроинформатика-99*. М.: МИФИ. Ч. 2. С. 18—24.

11. **Шевченко П. А., Фомин Д. В., Черников В. М., Виксне П. Е.** Применение микропроцессора NM6403 для эмуляции нейронных сетей // *Нейрокомпьютеры и их применение-99*. М.: ИПУ РАН, 1999. С. 81—90.

12. **Lindsey C., Lindblat T.** Survey of neural network hardware. SPIE. V. 2492. P. 1194—1205.

13. **Tahir K. Shah** Automata, neural networks and parallel machines: some emerging principles // World Scientific. 1999.

14. <http://neurnews.iu4.bmstu.ru>

15. **Цыганков В. Д.** Нейрокомпьютер и его применение. М.: Сол Систем. 1993.

16. **Аряшев С. И., Бобков С. Г., Сидоров Е. А.** Параллельный перепрограммируемый вычислитель для систем обработки информационных сигналов // *Нейроинформатика-99*. М.: МИФИ. Ч. 2. С. 25—33.

17. **Борисов В. Л., Капитанов В. Д.** Методика быстрого создания нейроускорителей // *Нейрокомпьютеры: разработка и применение*, 2000. № 1. С. 12—24.

18. **Роберт Хехт-Нильсен.** Нейрокомпьютинг: история, состояние, перспективы // *Открытые системы*. 1998. № 4.

---

Ордена Трудового Красного Знамени издательство "МАШИНОСТРОЕНИЕ"  
107076, Москва, Стромьинский пер., 4

---

Телефон редакции журнала (095) 269-5510

---

Художественный редактор *Т.Н. Погорелова*. Художник *В.Н. Погорелов*.  
Технический редактор *И.С. Павлова*. Корректор *Л.А. Богомолова*

Сдано в набор 10.07.2000. Подписано в печать 22.08.2000. Формат 60×88 1/8. Бумага офсетная. Печать офсетная.  
Усл.-печ. л. 2,94. Уч.-изд. л. 4,6. Заказ 973.

Журнал зарегистрирован в Комитете Российской Федерации по печати. Свидетельство о регистрации № 013895 от 05.05.98.

Отпечатано в Подольской типографии Чеховского полиграфического комбината  
142100, г. Подольск, ул. Кирова, 25

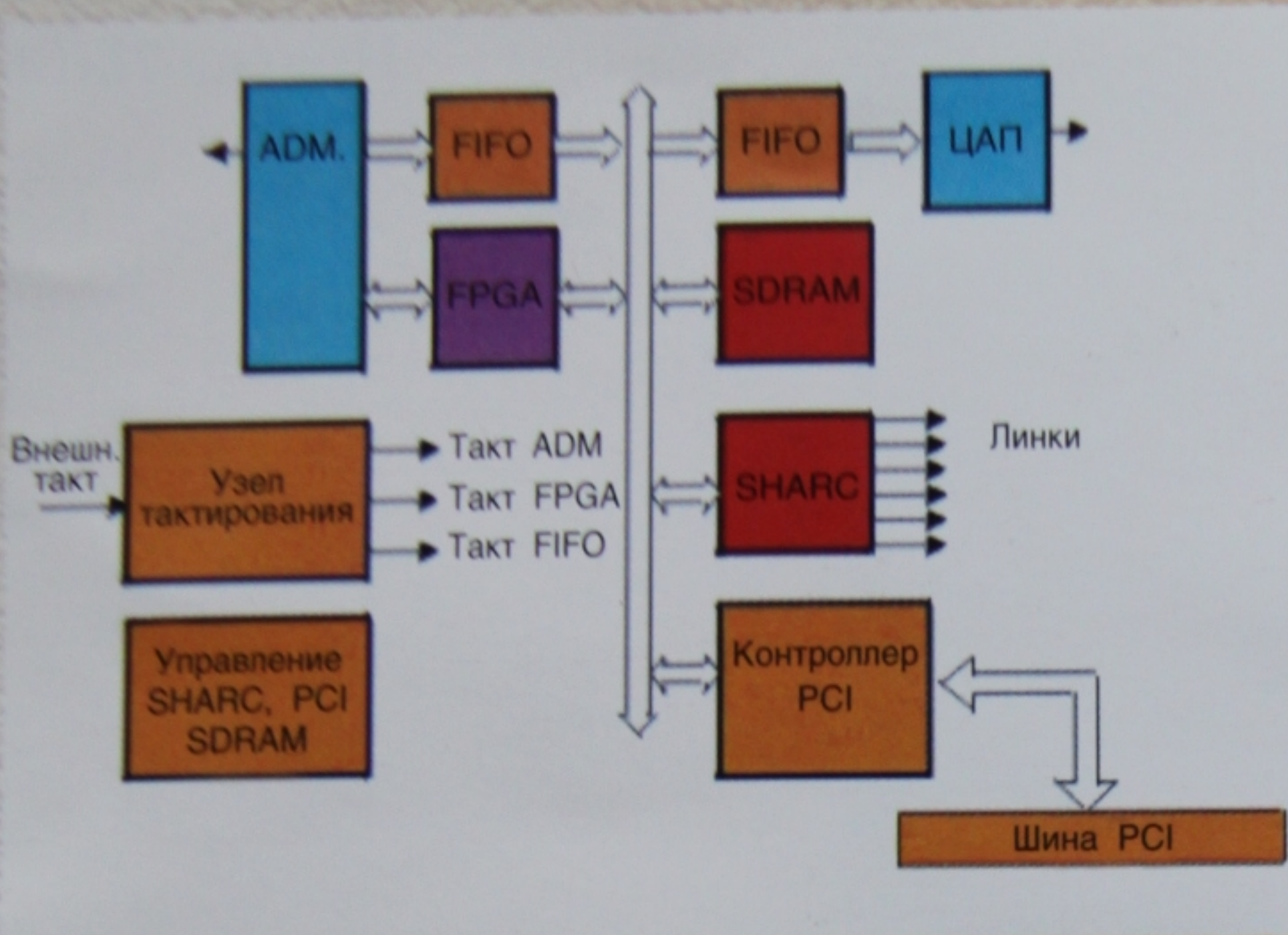


Рис. 17. Структурная схема платы ADP160PCI

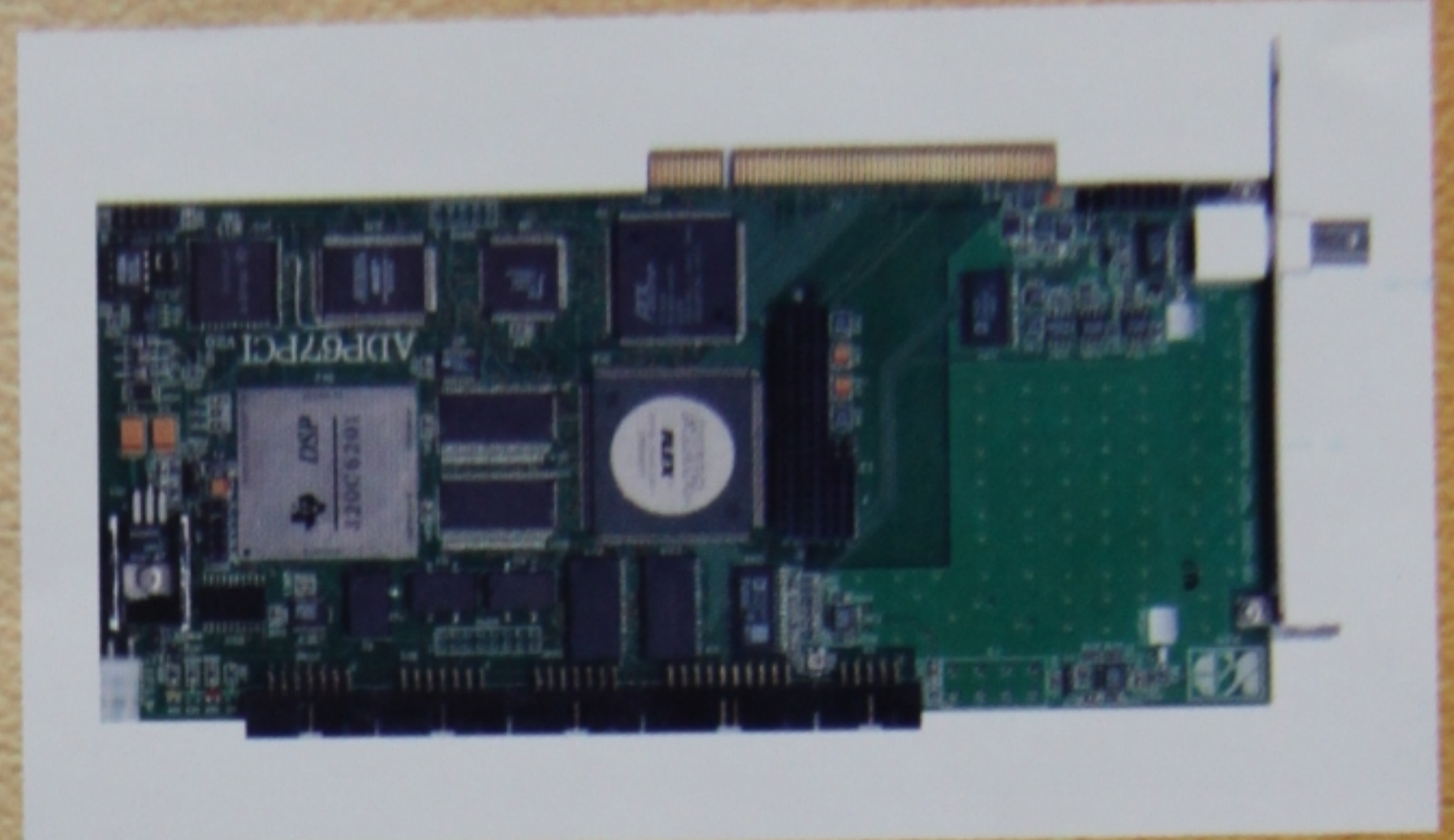


Рис. 19. Общий вид процессорного модуля на базе TMS320C6701

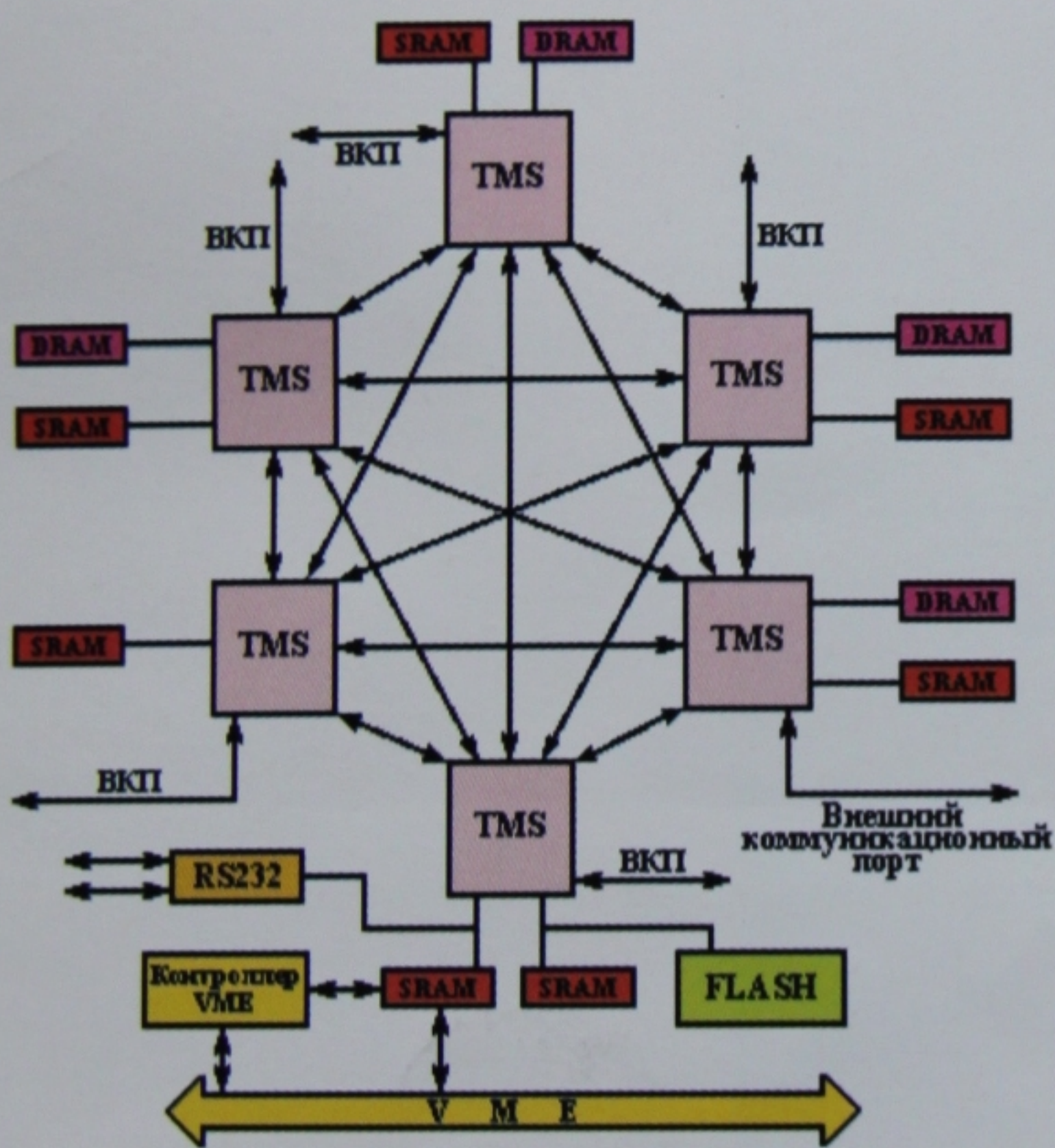


Рис. 20. Структурная схема платы M2

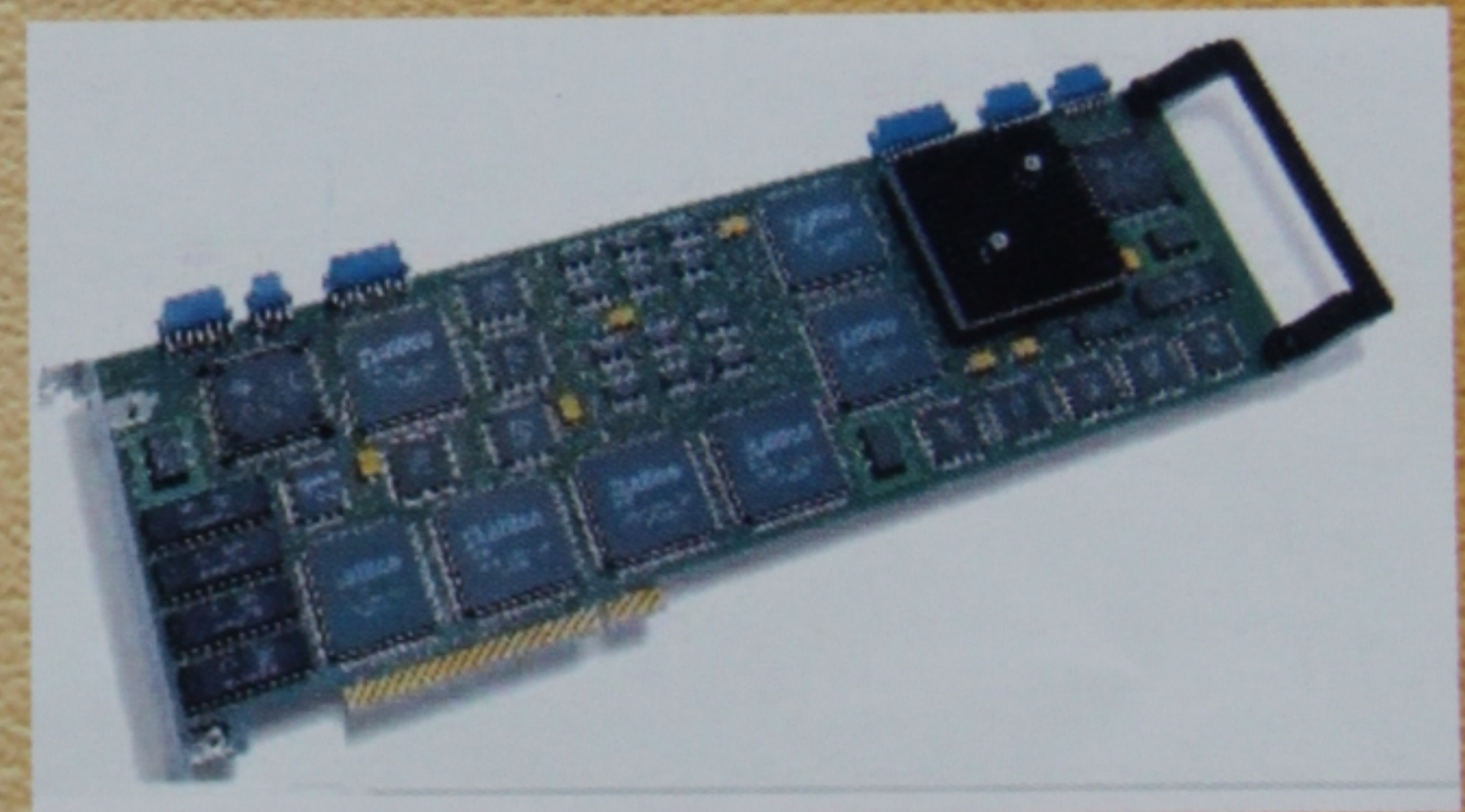


Рис. 22. Внешний вид "Synapse 2"

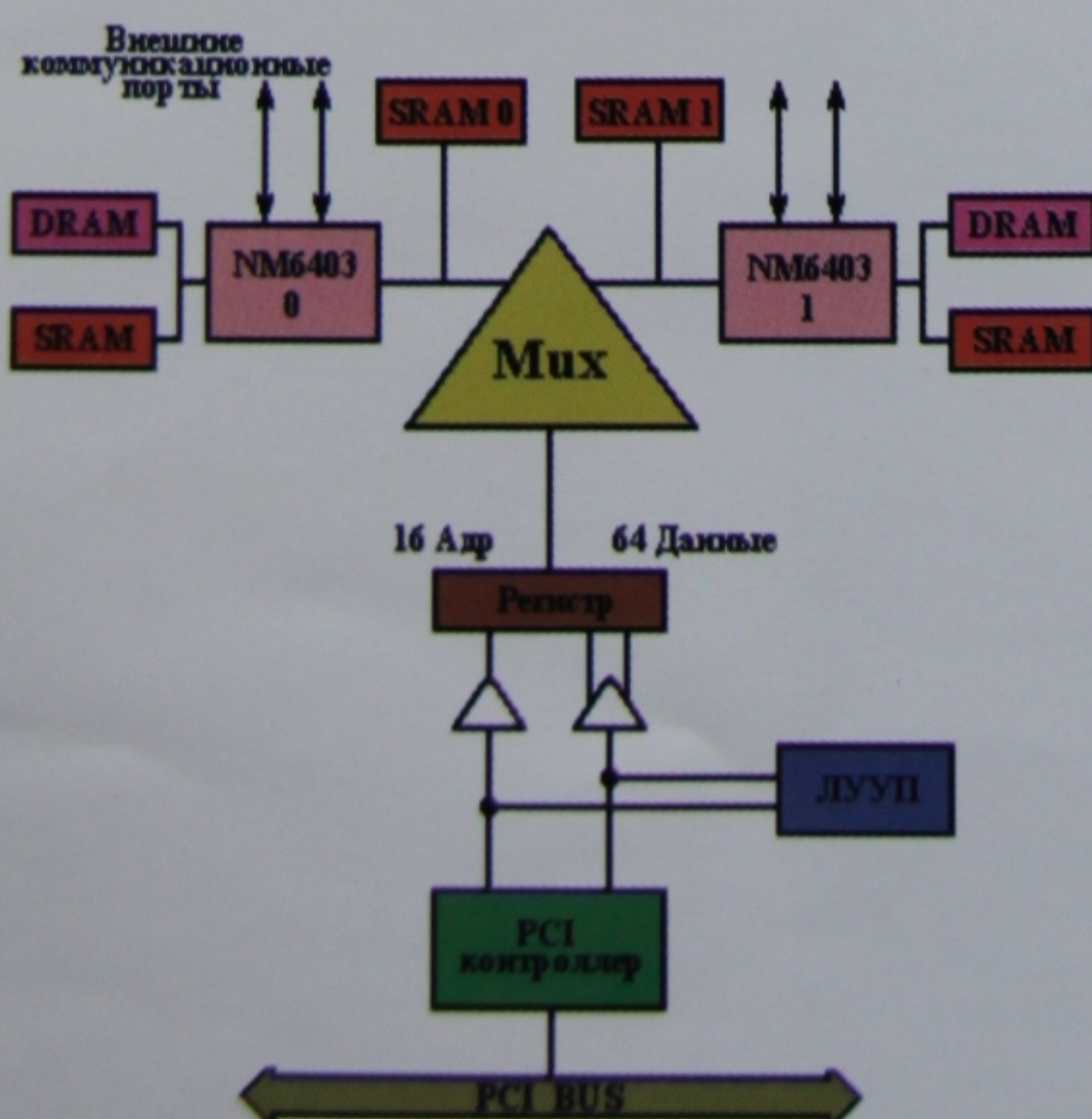


Рис. 21. Структурная схема модуля МЦ4.01

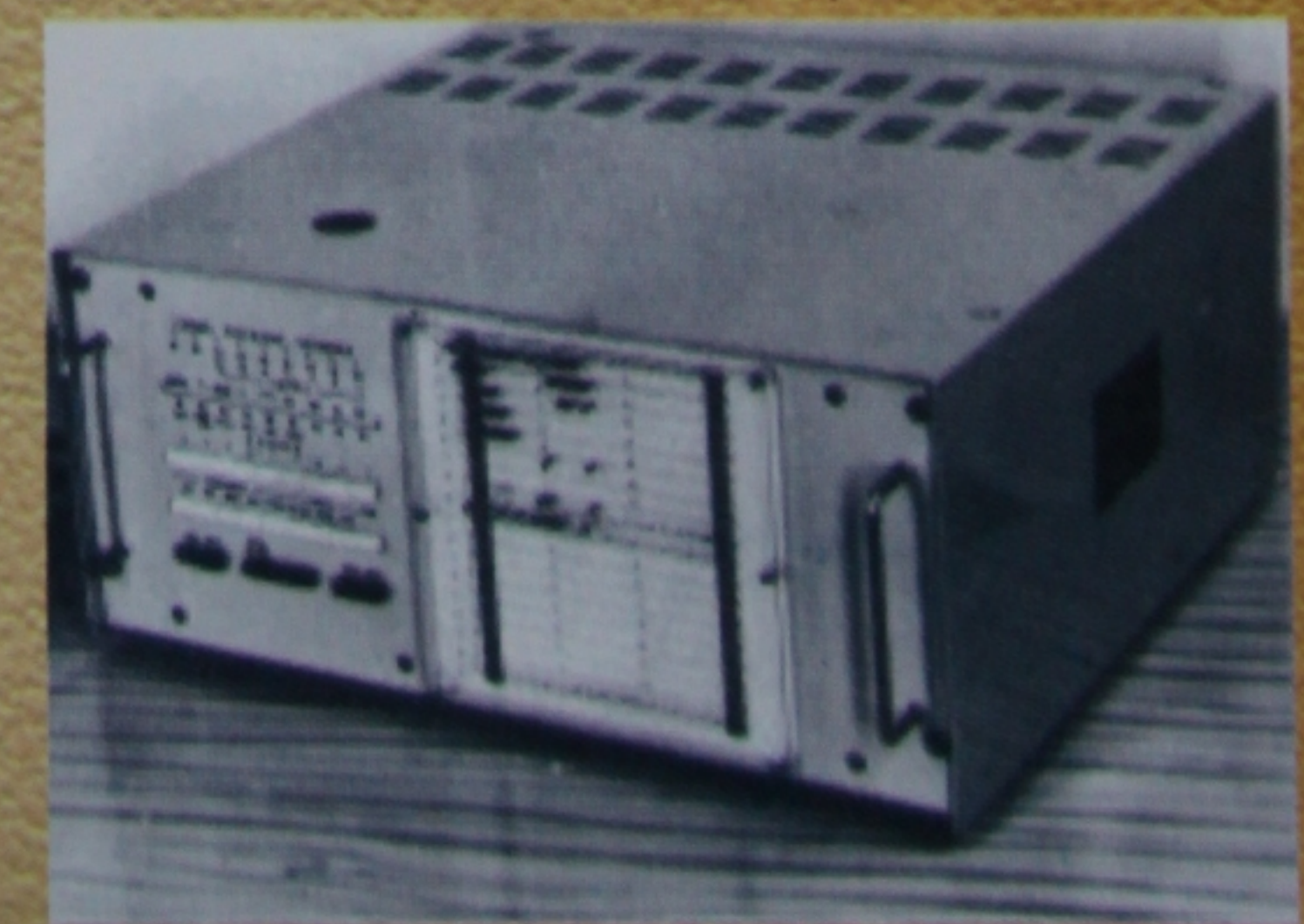


Рис. 25. Общий вид различных модификаций нейрокомпьютера "ЭМБРИОН"





Рис. 6. Принципы построения архитектуры TigerSHARC



Рис. 9. Структура NeuroMatrix NM6403

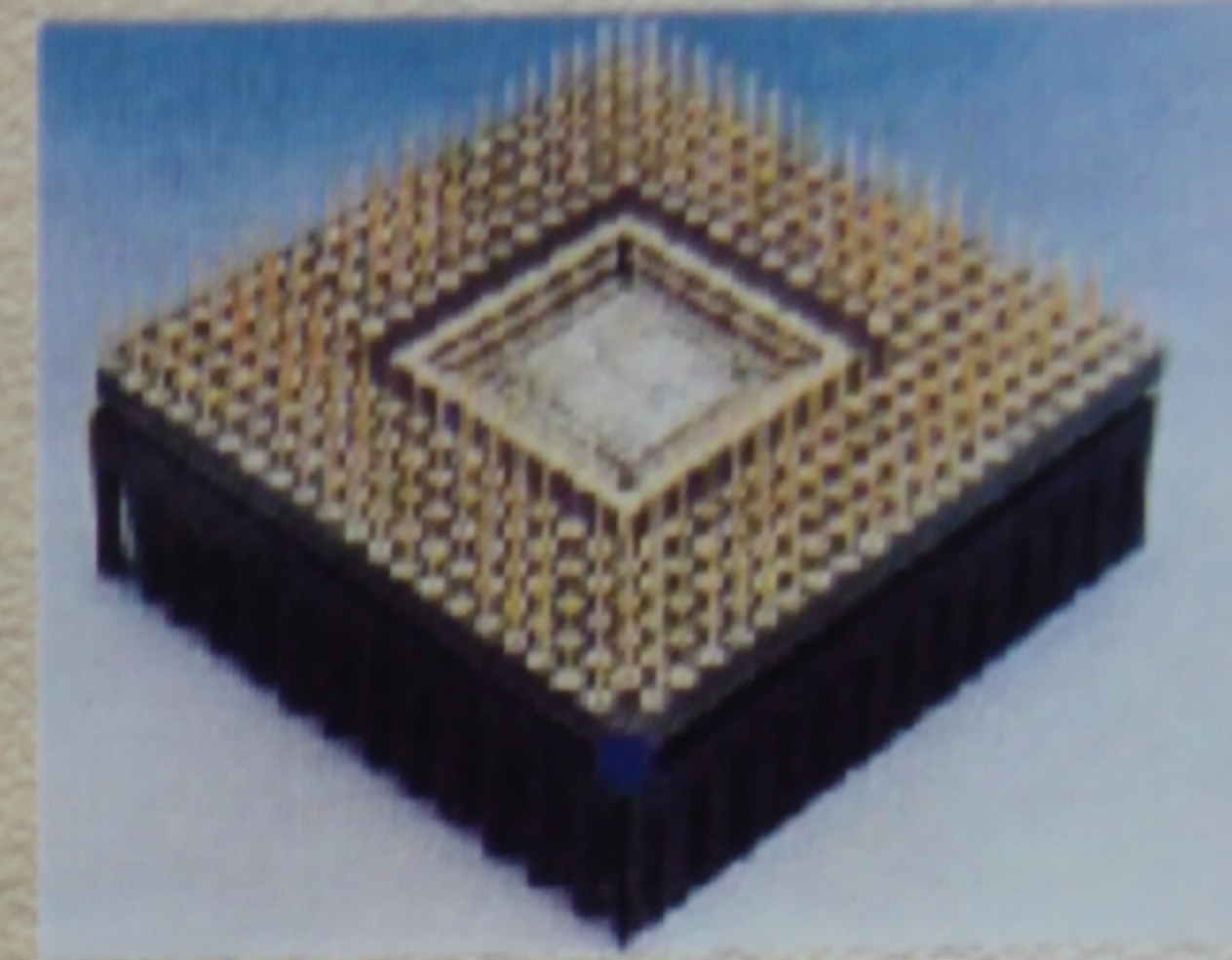


Рис. 10. Нейрочип MA16

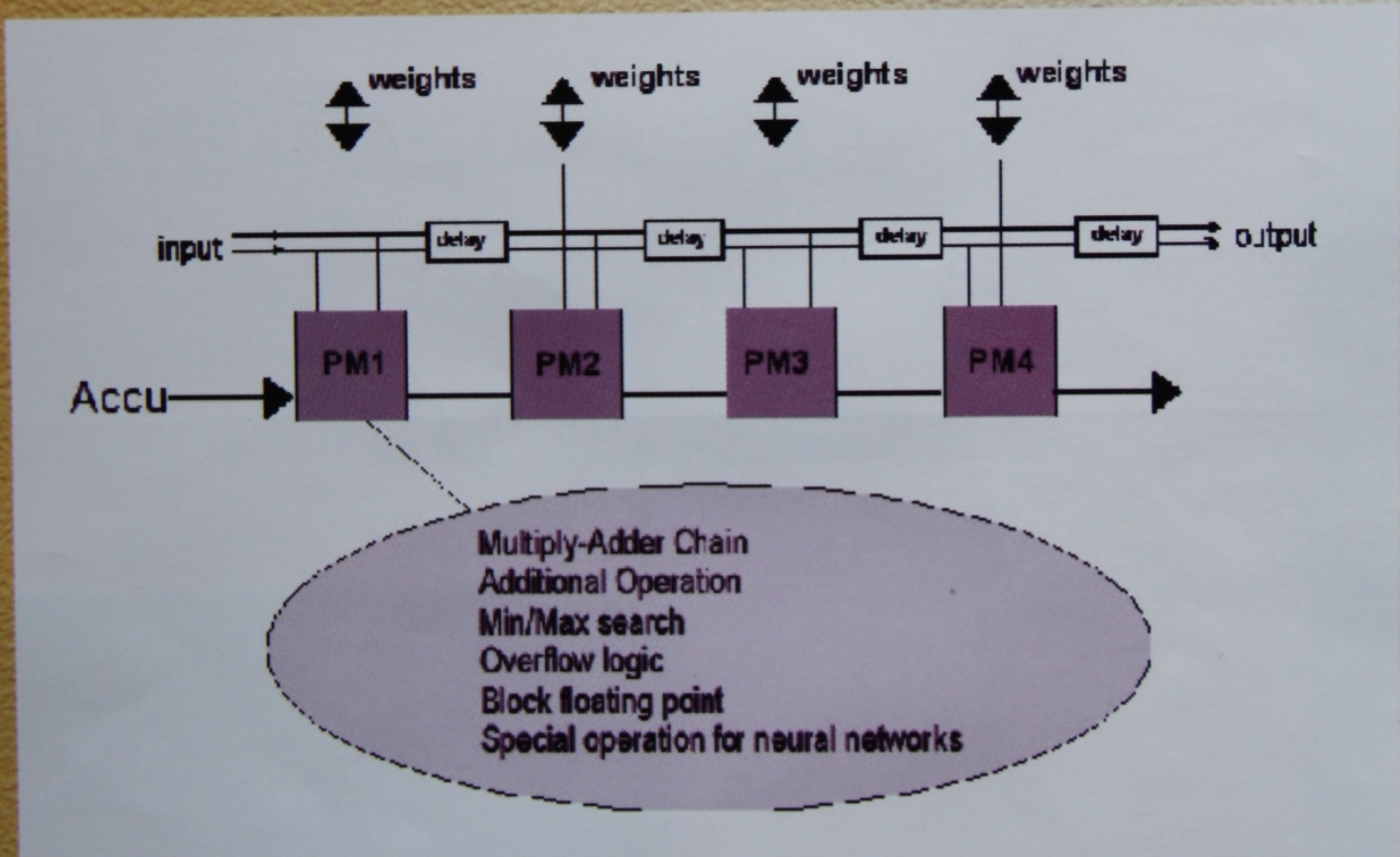


Рис. 11. Функциональная схема MA16

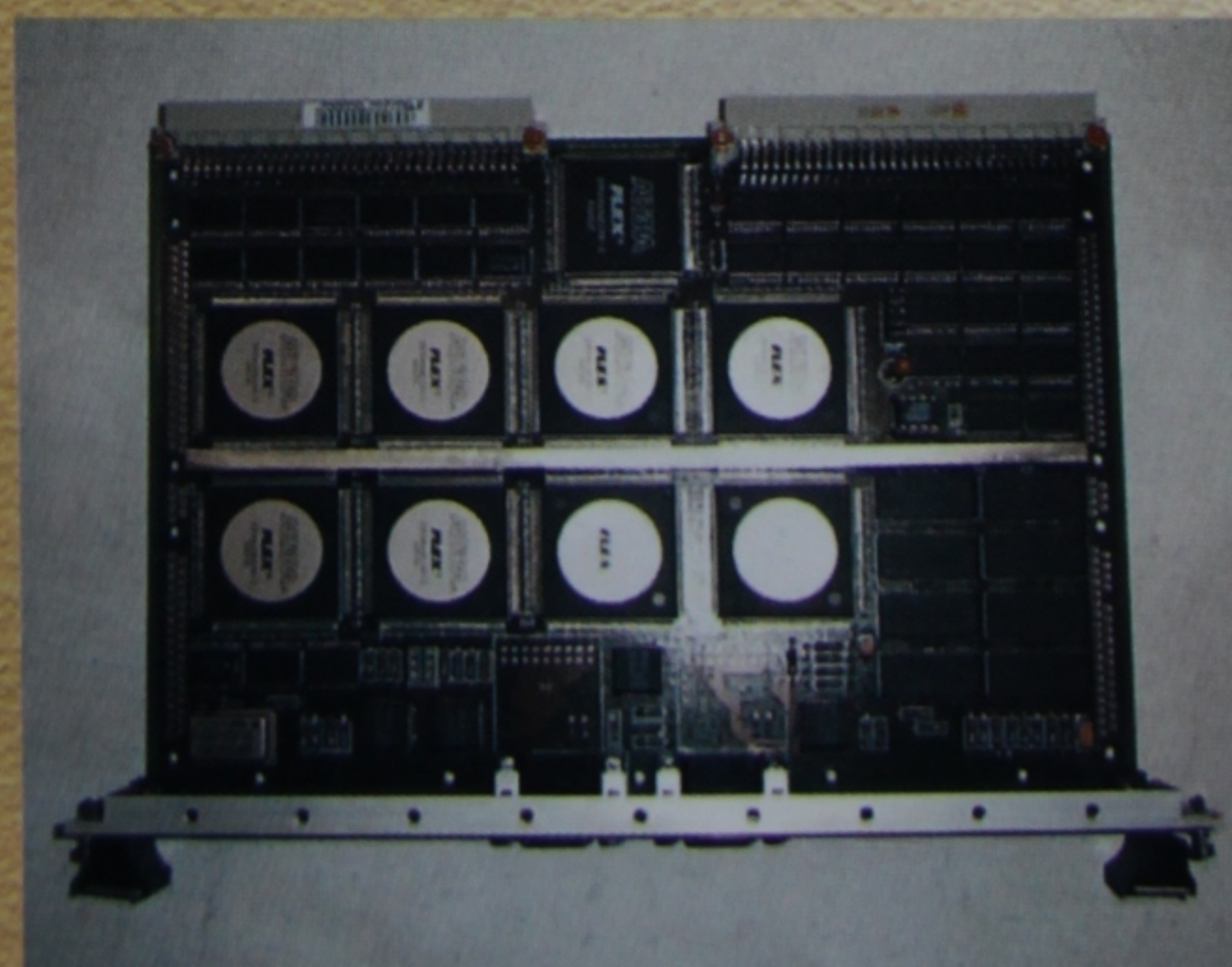


Рис. 14. Внешний вид ППВ