

УДК 681.31; 518.3

НЕЙРОЧИПЫ, НЕЙРОКОМПЬЮТЕРЫ И ИХ
ПРИМЕНЕНИЕ В ЭКСПЕРИМЕНТАЛЬНОЙ
ФИЗИКЕ ВЫСОКИХ И СВЕРХВЫСОКИХ ЭНЕРГИЙ

Н.М.Никитюк

Объединенный институт ядерных исследований, Дубна

ВВЕДЕНИЕ	431
НЕЙРОН КАК ЛОГИЧЕСКИЙ ЭЛЕМЕНТ. ПРОСТЫЕ КЛАССИФИКАТОРЫ	433
ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ. НЕЙРОПАКЕТЫ	438
ПРОЦЕСС ОБУЧЕНИЯ ИНС	442
АНАЛОГОВЫЕ И АНАЛОГОВО-ЦИФРОВЫЕ НЕЙРОЧИПЫ	443
НЕЙРОПЛАТЫ И НЕЙРОКРЕЙТЫ	447
ПРИМЕНЕНИЕ АНАЛОГОВЫХ И АНАЛОГОВО-ЦИФРОВЫХ НЕЙРОЧИПОВ В ЭКСПЕРИМЕНТАХ ПО ФИЗИКЕ ВЫСОКИХ ЭНЕРГИЙ	448
ПРИМЕНЕНИЕ АНАЛОГОВЫХ ИНС ДЛЯ ИЗМЕРЕНИЯ ФИЗИЧЕСКИХ ПАРАМЕТРОВ И УПРАВЛЕНИЯ ОБЪЕКТАМИ	458
ПРИМЕНЕНИЕ ТРАНСПЬЮТЕРОВ И ЦИФРОВЫХ СИГНАЛЬНЫХ ПРОЦЕССОРОВ	460
ЦИФРОВЫЕ НЕЙРОЧИПЫ И НЕЙРОКОМПЬЮТЕРЫ	463
ЗАКЛЮЧЕНИЕ	476
СПИСОК ЛИТЕРАТУРЫ	476

УДК 681.31; 518.3

НЕЙРОЧИПЫ, НЕЙРОКОМПЬЮТЕРЫ И ИХ ПРИМЕНЕНИЕ В ЭКСПЕРИМЕНТАЛЬНОЙ ФИЗИКЕ ВЫСОКИХ И СВЕРХВЫСОКИХ ЭНЕРГИЙ

Н.М.Никитюк

Объединенный институт ядерных исследований, Дубна

Рассмотрены архитектурные особенности и параметры ряда нейрочипов и нейрокомпьютеров, применяемых в научном приборостроении, описана тенденция их развития и использования в экспериментах по физике высоких и сверхвысоких энергий. Приведены сравнительные данные, характеризующие эффективность применения нейрочипов для отбора полезных событий, классификации элементарных частиц, реконструкции треков заряженных частиц и для поиска гипотетических частиц Хиггса, а также характеристики нейрочипов и ускорительных нейроплат отечественного производства.

Architecture peculiarity and characteristics series neural chips and neural computers used in scientific instruments are considered. Tendency development and its use in high energy and superhigh energy physics experiment are described. Compare data which characterize the efficient use neural chips for usefull event selection, classification elementary particles, reconstruction tracks charged particles and for serch hypotesis Higgs particles are given. The characteristics nativ neural chips and accelerated neural boards are considered.

1. ВВЕДЕНИЕ

Человеческий мозг, содержащий порядка 100 млрд нейронов, где каждый нейрон соединен с 1000 аналогичных ячеек, выполняющих относительно простые операции, по существу, является совершенным параллельным и, скорее всего, чисто аналоговым процессором. Самый совершенный компьютер трудно сравнить с человеческим мозгом отнюдь не только потому, что он содержит гигантское количество логических ячеек-нейронов. Принципиальное отличие прежде всего состоит в способе обработки информации, высокой степени параллелизма и связности между нейронами. Кроме того, данные в нейронных сетях не содержатся в отдельном блоке памяти, как в обычной ЭВМ. Они распределены, запоминаются и содержатся в виде весовых коэффициентов. Важно также, что в нейронных сетях операции вычисления и коммутации данных совмещены.

В типичной ЭВМ дискретного действия для обработки простейшего изображения требуется выполнить множество операций над числовыми данными,

получаемыми к тому же после преобразования естественных аналоговых сигналов в дискретные эквиваленты. После такого преобразования, как правило, теряется много первичной информации, и поэтому компьютер восстанавливает ее в приближенном виде. Другими словами, при обработке образов цифровая ЭВМ имеет крайне низкий КПД. В человеческом мозге (в первом приближении) реализуется «картинная логика», когда без пространственного или временного сканирования обработка данных выполняется путем наложения отдельных картин друг на друга. Отсюда высокая скорость обработки образов при довольно больших задержках в отдельном нейроне (несколько миллисекунд). Человеческий мозг может работать на уровне значительных помех. Мы сталкиваемся с удивительными фактами восприятия информации, поступающей от органов чувств, на уровне больших помех или в искаженном виде: из отдельных фрагментов картины восстанавливается целый образ, распознаются знакомые голоса на уровне уличных помех, а дети уже в раннем возрасте узнают лица своих близких и т.д. И, наконец, последнее и не менее удивительное: мозг способен обучаться без каких-либо побуждений извне.

В зарубежной литературе вопросам применения искусственных нейронных сетей в различных областях науки и техники уделяется серьезное внимание, о чем свидетельствуют многочисленные публикации. Здесь можно отметить популярные статьи [1–3]. В [1] рассматривается проблема психики и мозга. Отмечается разнообразие нервных клеток, которое отметил один из основателей современной науки о мозге Сантьяго Рамон-и-Кахаль. Описывается упрощенная схема работы нейрона. Работа [2] посвящена вопросам обучения нейронных сетей. Описывается суть широко используемого для настройки искусственных нейронных сетей (ИНС) алгоритма обратного функционирования. Ряд конкретных схем ИНС приведен в [3], где показано, каким образом решается проблема четности с помощью ИНС.

Уже сложилось несколько направлений в создании ИНС в интегральном исполнении. Появился ряд периодических изданий, посвященных нейросетям, например «Neural Networks», «IEEE Transaction on Neural Networks», «Neural Computation» и др. Среди популярных работ, опубликованных в конце 80-х годов, можно отметить обзор Т. Кохонена [4], в котором данная проблема рассматривается с исторической точки зрения и детально описываются отличия в работе биологического мозга от работы цифровых ЭВМ. Заслуживает внимания также книга упомянутого автора, изданная на русском языке [5]. Ряд сведений о нейросетях даны в обзоре [6]. Много полезных сведений о вычислениях с помощью ИНС и о классификаторах, применяемых в научных исследованиях с помощью ИНС, приводится в обзоре [7]. В [8] дается систематическое изложение основ нейровычислений и приводится ряд примеров использования ИНС в научных исследованиях и в измерительной технике. Применение нейрокомпьютеров в разработках военной техники США рассмотрено в обстоятельных обзорах [9,10]. В [11] описаны тео-

ретические основы различных моделей ИНС, дан обзор работ, посвященных использованию программных ИНС в экспериментах по физике высоких энергий. Из публикаций последних лет следует отметить тематический выпуск, посвященный нейрокомпьютерам и их применению, где, в частности, рассмотрены вопросы концептуального развития нейрокомпьютеров [12], описаны программные продукты (нейропакеты) [13], приведены параметры популярного нейрочипа ETANN [14] и оптических нейрокомпьютеров [15] и т.д. Можно отметить три основных сложившихся направления в исследованиях ИНС: математическое, физическое и технологическое.

Цель данного обзора состоит в том, чтобы представить тенденцию в развитии архитектуры современных нейрочипов и их применение в научных исследованиях. Обзор состоит из десяти разделов, заключения и приложения. Во введении рассматривается актуальность проблемы и отмечаются принципиальные различия между современными суперкомпьютерами и нейрокомпьютерами. Во втором разделе обзора описывается структура нейрона как логического элемента. Приводится приближенная схема биологического нейрона. Третий раздел посвящен краткому обзору наиболее известных моделей искусственных нейронных сетей (ИНС). Методам обучения ИНС посвящен четвертый раздел обзора. В пятом разделе описываются интегральные аналоговые и аналогово-цифровые нейрочипы. Шестой раздел посвящен описанию нейроплат и нейрокрейтов. В седьмом разделе обзора рассматриваются примеры применения популярного аналогового нейрочипа ETANN, а также других аналоговых нейрочипов. В восьмом разделе приводятся примеры применения аналоговых нейрочипов для измерения физических параметров и управления объектами. Применение для построения ИНС транспьютеров и цифровых сигнальных процессоров рассмотрено в девятом разделе. В десятом разделе описана архитектура и приведены примеры применения цифровых нейрочипов и нейрокомпьютеров. В приложении дана сводная таблица рассматриваемых нейрочипов и нейрокомпьютеров.

2. НЕЙРОН КАК ЛОГИЧЕСКИЙ ЭЛЕМЕНТ. ПРОСТЫЕ КЛАССИФИКАТОРЫ

Кратко рассмотрим то, что известно о нейроне как логическом элементе. Исходя из современных представлений, биологический нейрон можно схематически изобразить, не претендуя на строгость, как это показано на рис. 1. Нейрон состоит из тела, называемого сомой (soma), дендритов, по которым поступают входные сигналы, и аксона. С помощью аксонов сигналы от одного нейрона передаются к другим нейронам через синапсы, которые находятся на стыке между аксонами и дендридами. Предполагается, что мозг содержит до 10^{14} синапсов. Каждому синапсу ассоциируется сила связи или

весовой коэффициент. Возбужденные нейроны посылают электрические импульсы, которые исходят из тела нейрона. Сигналы имеют амплитуду около

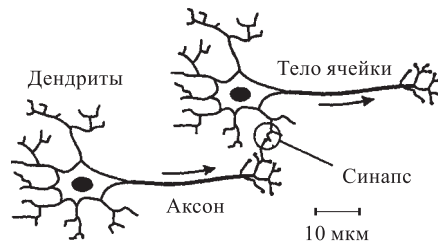


Рис. 1. Упрощенная схема двух биологических нейронов. Дендриты являются входами нейронов. После возбуждения нейрона сигнал распространяется вдоль аксона в направлении, указанном стрелкой. Взаимодействие нейронов происходит на стыке, который называется синапсом [8]

искусственного (формального) нейрона. Стык или узел между нейронами называется синапсом. Он выполняет преобразование входной информации в сигналы, воспринимаемые нейроном. С точки зрения логики нейрон представляет собой в первом приближении пороговый элемент, который с помощью синапсов собирает сигналы и суммирует их в аналоговом сумматоре. Каждый нейрон время от времени оценивает эту сумму по отношению к пороговому уровню, после чего он оставляет свой выходной сигнал без изменения либо изменяет его. Если сумма превышает определенный порог, то выходной сигнал распространяется по аксонам дальше, причем сигнал может проходить через нейрон только в одном направлении. Искусственный нейрон имеет определенное количество входов с весами w_{ij} , сумматор с суммирующей функцией, пороговый элемент и чаще всего один выход. Связи между нейронами имеют резистивный характер. Для реализации пороговой функции в нейрон вводится смещение или смещающие элементы, выходы которых соединены со входами суммирующего элемента. На рис. 2 приведена схема идеализированного искусственного нейрона [2]. В общем виде выход нейрона x_j в однонаправленной сети, содержащей n нейронов (т.е. сети без обратных связей с выходов на входы), описывается следующим уравнением:

$$x_j = f \left(\sum_{i=1}^n w_{ij} v_i + s \right), \quad (1)$$

где w_{ij} — вес связи между нейронами j и i , v_i — входной сигнал нейрона i , s — порог (смещение), который складывается со значением взвешенной

100 мВ длительностью 1 мс. Каждый нейрон имеет тысячи связей с другими нейронами. Обучение происходит путем изменения силы связи синапсов. В результате меняется влияние одного нейрона на другой. Причем, в зависимости от состояния синапса, он или возбуждает следующий нейрон, или запрещает его срабатывание. Обнаружено различие между клетками с короткими аксонами, которые взаимодействуют с соседними нейронами, и клетками с длинными аксонами, от которых сигналы идут в другие участки мозга.

Рассмотрим электрическую схему

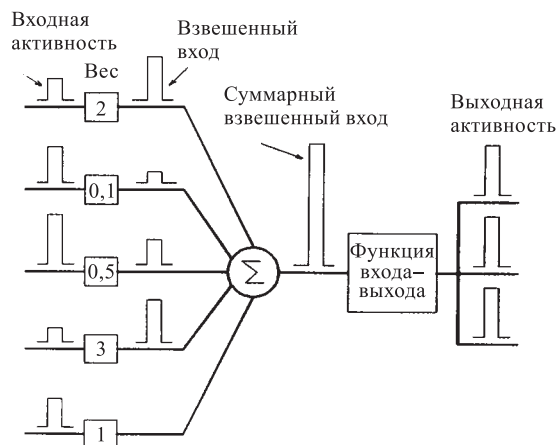


Рис. 2. Блок-схема формального (искусственного) нейрона с пятью входами. Нейрон обрабатывает электрические сигналы, поступающие на его входы. Каждый входной сигнал умножается на весовой коэффициент. С помощью сумматора значения взвешенных сигналов складываются. Если суммарный сигнал превышает порог, который вычисляется с помощью функции входа-выхода, на выходе нейрона появляется импульс

суммы, f — функция активации (например, сигмоидная функция, пороговая функция, линейная функция и др.). Используемая сигмоидная функция часто определяется из уравнения

$$y = 1/(1 + e^{-x}).$$

Первая часть выражения (1) представляет собой процедуру матричного умножения матрицы весов W и вектора активности (входного воздействия) v и является наиболее трудоемким процессом с точки зрения вычислений:

$$x = f(Wv).$$

Сигмоидная функция (рис. 3) дает такой же выход, как и обычная пороговая функция, но математически она проще. Однако, в отличие от пороговой функции, сигмоидная функция имеет производную, что весьма удобно в процессе обучения тех сетей, в которых используются алгоритмы, требующие выполнения такой операции. Следует отметить, что перемножение вектора на матрицу часто используется в численном анализе, поэтому по аналогии с созданными копроцессорами с плавающей точкой появились нейропроцессоры, выполняемые на стандартных печатных платах с

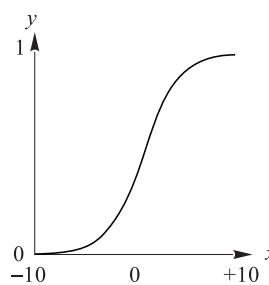


Рис. 3. Сигмоидная функция, задаваемая уравнением $y = 1/(1 + e^{-x})$

интерфейсом для связи с основной ЭВМ и возможностью каскадирования для повышения производительности.

Для моделирования ИНС искусственные нейроны группируются в линейные матрицы, называемые слоями или уровнями. Для решения различного рода задач, которые связаны с распознаванием образов, классификации объектов и др., создаются искусственные нейронные сети, которые представляют собой физические модели, состоящие из множества искусственных нейронов. Из многих десятков известных моделей нейронных сетей наиболее широкое применение нашли три модели: модель Хопфильда [16], машина Больцмана [17] и модель, описываемая радиально-базовыми функциями [18]. В процессе описания функционирования нейронных сетей будут кратко рассмотрены и другие используемые модели.

Наиболее широкое применение в экспериментах по физике высоких энергий нашли первые две модели. Машины Больцмана представляют собой класс сетей, в которых используется шумовой процесс для нахождения

глобального минимума оценочной функции.

Модель Хопфильда базируется на представлении нейрона в виде сумматора-дискриминатора, на входах которого имеются весовые резисторы, моделирующие функции синапса. Элементу модели Хопфильда соответствует схема, изображенная на рис. 2.

При поступлении сигнала на вход искусственного нейрона он умножается на вес соответствующей связи и нейрон складывает все взвешенные входные значения. Действие такого элемента во многом зависит от того, какую функцию он выполняет: линейную, пороговую или сигмоидную. Сигмоидальные элементы можно рассматривать как приближения к реальным нейронам.

Путем изменения порога суммирования (смещающего элемента) можно настроить одну и ту же схему на выполнение операций конъюнкции или дизъюнкции, что следует

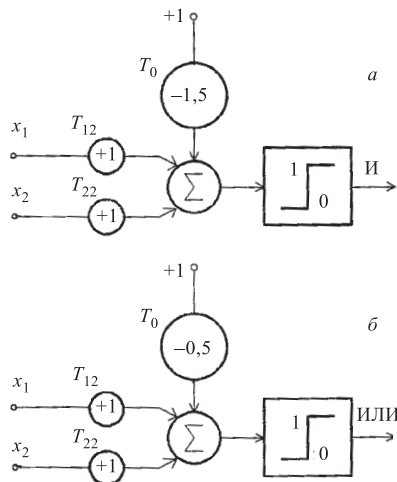


Рис. 4. Логические операции двух переменных И (а) и ИЛИ (б), выполняемые искусственным нейроном в зависимости от величины порога: T_{12} , T_{22} — веса; T_0 — смещение

из рис. 4. Здесь мы имеем пример элементарной однослойной сети. В экспериментах по физике высоких энергий широко используется классификация элементарных частиц по определенным признакам. Это может быть тип событий, их особенность и проч. Связи между событиями могут быть линейными

и нелинейными. Так, линейная дискриминация для одномерной физической переменной эквивалентна пороговой функции, как это показано на рис. 5 [19]. Здесь мы имеем пример линейного классификатора, который можно реализовать с помощью нейрона с одним входом и со смещающим элементом, на выходе которого всегда присутствует сигнал логической единицы.

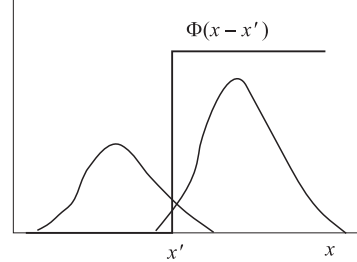


Рис. 5. Пороговая функция $\Phi(x-x')$, используемая для фильтрации событий одной переменной, которая эквивалентна линейной дискриминации для одномерной физической переменной

Для большого числа переменных линейная дискриминация возможна, если объекты (категории) линейно различимы. В таком случае решение возможно с помощью простой нейронной сети, реализующей дискриминатор Фишера, известный из статистических методов обработки данных [20]. В [21] кроме свойств дискриминатора Фишера описывается также ряд других классификаторов (алгоритм обратного функционирования, классификатор Гаусса и проч.), применяемых в экспериментах по физике высоких энергий.

Рассмотрим кратко суть дискриминатора Фишера. Допустим, что мы хотим разделить два класса событий A и B , содержащих события с двумя измененными переменными s_1 и s_2 (рис. 6). Видно,

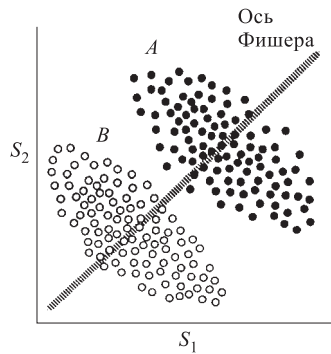


Рис. 6. Пример классификации групп событий A и B с помощью дискриминатора Фишера. S_1 и S_2 — оси координат

что если мы попытаемся дискриминировать события этих классов путем использования одномерного линейного классификатора (проецируя их координаты на оси S_1 или S_2), то окажется, что значительная часть событий наложится на другие. Однако, если проецирование выполнить на ось Фишера, то в значительной степени эти два класса событий будут различимы. Проекция на ось Фишера [22] описывается уравнением (1), которое реализуется с помощью простого типа однонаправленной нейронной сети, имеющей входные и выходные нейроны со смещением. Далее, если граница между классами событий A и B носит нелинейный характер, то проблема решается

путем введения еще одного, скрытого (дополнительного) слоя, который чаще всего подобен входному слою, но у него нет внешних выводов. В результате получают широко распространенные трехслойные ИНС, которые состоят из трех групп элементов: слоя входных элементов, на которые поступают

входные сигналы, слоя скрытых элементов, входы которых определенным образом соединены с выходами входных элементов. В свою очередь, входы третьего выходного слоя подключены к выходам скрытого слоя.

3. ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ. НЕЙРОПАКЕТЫ

Искусственные нейронные сети являются мощным инструментом для решения таких проблем, как распознавание образов, анализ данных и нелинейный контроль. Особенно эффективно с помощью ИНС решается совокупность задач, возникающих при распознавании образов, относящихся к категории, которую называют нерегулярными (случайными) задачами. Широкие перспективы открываются для использования ИНС в научном приборостроении, где требуется быстрое решение нелинейных задач. Формально нейронная сеть определяется как сеть с конечным числом слоев из однотипных нейронов. Популярными оказались две концепции построения ИНС: многослойные однонаправленные сети и сети Хопфильда, которые рассматриваются как обобщение однонаправленных сетей в силу нерегулярности связей между элементами и наличия обратной связи. В однонаправленной сети или в сети с прямыми связями потоки информации могут проходить только в направлении от предыдущего слоя к последующему.

В качестве примера ИНС первого типа может служить двухслойный персептрон Розенблата [23], у которого каждый элемент входного слоя соединен со входом каждого нейрона выходного слоя. Однако такая сеть имеет ограниченные функциональные возможности, поскольку не все комбинации входных значений правильно дешифруются. Поэтому широкое применение находят трехслойные сети. В такой сети входной вектор входного воздействия поступает на элементы входного слоя. Далее он перемножается на матрицы весовых коэффициентов этого уровня, после чего новая информация по связям поступает на средний (скрытый) уровень. Аналогичная операция повторяется на среднем и выходном уровнях. В настоящее время находят применение также многослойные сети, которые содержат несколько скрытых слоев. Вектор входного воздействия в однонаправленной сети распространяется от элементов входного слоя к скрытым элементам, образующим второй слой, которые передают сигналы к выходному слою. Скрытые элементы ведут себя оригинально: они могут создавать свои собственные представления о входных данных. Так, веса связи между входными и скрытыми элементами определяют, при каких сигналах на входах скрытый элемент будет активным. Сети со скрытыми слоями позволяют решать проблему четности, для реализации которой с помощью булевых выражений требуется инверсия переменных. На рис. 7 в качестве примера приведена прямоточная сеть, выполняющая функцию «ИСКЛЮЧАЮЩЕЕ ИЛИ» [3]. Числа 1,5; 0,5 и -2 на

линиях обозначают веса. Если на один входной элемент подается 1, на другой 0, то на выходе скрытого элемента будет 0, поскольку его порог превышает 1. Однако с учетом того, что входные элементы имеют еще связи с выходным элементом, последний все-таки возбуждается (полагаем, что сигналу нечетности соответствует логическая единица). Наличие двух единиц на входах сети возбуждает скрытый элемент, который блокирует активность двух входных элементов, и на выходе сети будет сигнал логического нуля.

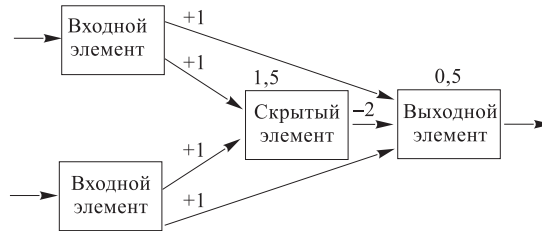


Рис. 7. Трехслойная однонаправленная сеть со скрытым элементом, выполняющая функцию «ИСКЛЮЧАЮЩЕЕ ИЛИ». Выходной элемент срабатывает только тогда, когда на один из входных нейронов подается сигнал, соответствующий логической единице. Числа, проставленные на соединениях, — это весовые коэффициенты (веса связи), а находящиеся над элементами — величины смещения

Вообще говоря, доступ к скрытому слою — это скорее всего технологическая проблема, поскольку стоимость нейрочипа во многом зависит от количества контактов в корпусе.

В общем виде однонаправленная трехслойная сеть содержит n входных нейронов, m скрытых нейронов и k выходных нейронов. Исходя из этого такую сеть мы будем называть $n - m - k$ -сетью.

Технологическая база ИНС. В современных суперЭВМ реализация нейронных сетей и связанных с ними идей выполняется программным или аппаратным способом. При аппаратной реализации в основном используются три подхода: аналоговый, цифровой и гибридный. Аналоговые ИНС имеют определенные преимущества при создании ИНС невысокой степени интеграции. Суть проблемы состоит в том, что для создания аналоговых ИНС требуется изготовление высокоточных резисторов и конденсаторов. Кроме того, они подвержены воздействию электрических помех и шумов. И, наконец, требуется множество преобразователей уровней при стыковке с управляющими компьютерами.

В [15] отмечается, что переход от булевой к пороговой логике, используемой в ИНС, расширяет круг технологий изготовления нейрочипов и нейрокомпьютеров. Для изготовления нейрочипов используются в основном три типа технологий: технология электронных сверхбольших интегральных ми-

кросхем (СБИС), использование заказных СБИС или электрически перепрограммируемых логических матриц (ЭПЛМ) и оптические методы обработки данных.

Возможности системы автоматизированного проектирования (САПР) устройств на базе ИНС с использованием ЭПЛМ описаны в [24]. Ее особенности заключаются в регулярности структуры с выделенными трассами для соединения макроэлементов в требуемую функциональную схему, применении стандартных узлов, содержащихся в каталоге, возможности моделирования разработки с использованием программного комплекса.

Широкое применение, в том числе и для создания ИНС, получила САПР фирмы «Xilinx». Одна из последних разработок имеет 500 тысяч вентилях в одном корпусе микросхемы и работает на частоте 150 МГц. Заказной СБИС имеет статическую память емкостью 2 Мбит [25]. Однако специфика пороговой нейронной логики неадекватна технологии СБИС в силу того, что в существующих электронных схемах вся обработка информации происходит в двумерной плоскости СБИС или пластины, что создает трудности при реализации нейронных сетей. Нетрудно создать на одном кристалле большое количество искусственных нейронов. Суть проблемы заключается в том, что при числе нейронов N в полносвязной ИНС требуется N^2 синапсов. В результате более 90% площади кристалла используется под устройства, реализующие матрицу адаптивных синапсов. Кроме того, в полносвязной ИНС имеется множество пересечений проводников, каждое из которых требует дополнительного слоя металлизации и двух слоев изолятора. Поэтому разработка сложной ИНС, содержащей несколько сотен искусственных нейронов с использованием технологии СБИС, создает определенные трудности, которые можно было бы преодолеть, если бы в интегральной электронике можно было создавать трехмерные структуры на одном кристалле. На современном этапе создания нейронных сетей и нейрокомпьютеров весьма перспективной можно считать технологию компоновки трехмерных кремниевых пластин [15].

Наряду с электрическими моделями ИНС находят также применение оптические нейронные сети [15]. Основным преимуществом таких устройств является то, что световые лучи не влияют друг на друга и могут многократно перекрещиваться. Кроме того, с помощью оптических методов более просто обрабатываются образы в реальном времени. В указанной работе дается обзор и приводится обширный список литературы, посвященной вопросам применения оптических методов и нанотехники для аппаратной реализации ИНС, а также современная оптическая элементная база и задачи, решаемые оптическими нейрокомпьютерами.

В [26] описана конструкция оптоэлектронного нейрокомпьютера, центральным элементом которого является оптоэлектронный векторно-матричный умножитель. Разработка световых нейронных сетей предполагает использование модулированных лазерных лучей, дифракционных решеток и го-

логафических систем. Более подробное рассмотрение принципов действия оптических нейронно-сетевых компьютеров выходит за рамки данного обзора. Читателям, интересующимся данной проблемой, рекомендуем ознакомиться с работой [27].

Наряду с ИНС широкое применение в научном приборостроении находят также нейрокомпьютеры. «Нейрокомпьютер — это вычислительная система с архитектурой аппаратного и программного обеспечения, адекватной выполнению алгоритмов, представленных в нейросетевом логическом базисе» [12]. Однако в ряде статей вместо термина «нейрокомпьютер» используется более широкое понятие «нейронная сеть», хотя в таких сетях отсутствуют искусственные нейроны. Примеры таких сетей будут рассмотрены в седьмом разделе.

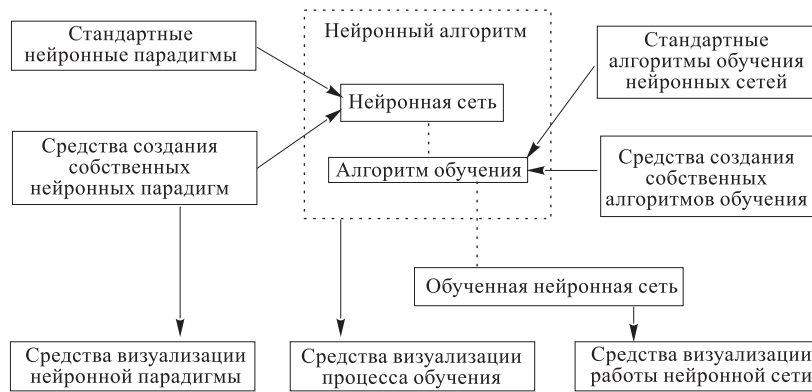


Рис. 8. Общая структура нейропакета

Нейропакеты. Как отмечается в [13], для реализации ИНС с помощью мощного парка современных вычислительных машин (в том числе и с использованием персональных компьютеров) необходимо иметь некоторый программный продукт, который моделирует нейрокомпьютер. Такая система программ получила название нейропакета. По существу, нейропакет является промежуточным звеном между архитектурой обычной ЭВМ и нейросетевым логическим базисом и аналогом широко известных компиляторов. Нейропакет является программной реализацией алгоритмов решения задач. Для решения различного типа задач уже создано 28 стандартных нейросетевых структур (парадигм). На рис.8 приведена общая структура нейропакета. В цитируемой выше работе приводится также классификация нейропакетов. Среди них можно отметить средства разработки (библиотеки запрограммированных нейронных парадигм и алгоритмов обучения), универсальные нейропакеты, предоставляющие пользователю работу с несколькими нейронными

парадигмами и имеющие средства для создания собственных разработок, специализированные нейропакеты и т.д. Таким образом, с помощью нейропакетов создаются виртуальные ИНС, для реализации которых могут быть использованы различные архитектуры параллельных процессоров: матричные процессоры, архитектуры типа «гиперкуб» и вычислители, создаваемые на базе современных транспьютеров и цифровых сигнальных процессоров.

4. ПРОЦЕСС ОБУЧЕНИЯ ИНС

Важным свойством ИНС является то, что они сами себя программируют в процессе обучения. Если рассматривать в общем, то ИНС обучаются путем ввода наблюдаемых образцов в модель. Далее они сравнивают входные образцы с хранящимися. Когда сеть анализирует и классифицирует хранящиеся образцы в сравнении с входными, в сети формируются новые весовые коэффициенты и осуществляется самообучение [28]. Таким образом, процесс обучения ИНС в основном заключается в вычислении весовых коэффициентов и установлении связей между нейронами. Поэтому количество устанавливаемых связей в секунду является одним из основных параметров ИНС не только в процессе обучения, но и в процессе эксплуатации. Обученная для решения определенного класса задач ИНС может принимать решения на базе знаний, полученных в процессе обучения на основе реальных образцов, несколько отличающихся от используемых для обучения. Как правило, весовые коэффициенты различных входов ИНС хранятся в виде матрицы, так что входной вектор умножается на указанную матрицу весовых коэффициентов. Поэтому обучение ИНС распознаванию заданного класса образов требует много машинного времени, поскольку количество сочетаний весовых коэффициентов весьма велико даже в простой сети.

Основы современного обучения сформулированы в правиле Хебба: «Если аксон клетки A достаточно близок к клетке B , чтобы возбуждать ее, и если он повторно или настойчиво принимает участие в ее срабатывании, то в одной из этих клеток или в обеих происходит некоторый процесс метаболического изменения, так что эффективность клетки A как одной из клеток, вызывающих срабатывание клетки B , увеличивается» [11]. Таким образом, путем изменения синаптических связей обеспечивается тренировка сети и хранение данных. Самый простой метод тренировки заключается в том, что на входах сети задается контрольный образец, например, изображение цифры, и сравнивается активность на выходных ячейках с желаемым поведением. Затем подсчитывается погрешность, определяемая как квадрат разности между действительным и желаемым выходом. С целью уменьшения погрешности вес связи уменьшается. Проблема заключается в том, что каждый вес необходимо изменять на величину, пропорциональную скорости, с

которой изменяется погрешность (производная погрешности по весу). Этот процесс требует много машинного времени, особенно для сложной многослойной сети. Более эффективным и широко используемым в физике высоких энергий и в других направлениях исследований является метод обратного функционирования (обратного распространения), предложенный Вербосом в 1974 г. и изложенный в доступной форме в [2]. Суть алгоритма заключается в следующем. Чтобы выполнить обучение сети, необходимо корректировать веса каждого нейрона таким образом, чтобы уменьшалась погрешность — разница между действительным и желаемым выходом. Для этого вычисляется производная от погрешности по весам. Затем в зависимости от величины и знака погрешности значения соответствующих весов меняются. Этот процесс повторяется до тех пор, пока не получится желаемый результат. Метод обратного функционирования позволяет эффективно настраивать многослойные сети, поскольку доступ к внутренним слоям, как правило, невозможен. Данный алгоритм успешно реализуется с помощью компьютеров. В последнее время в связи с появлением цифровых нейрочипов разработан ряд более простых и эффективных алгоритмов обучения ИНС, которые будут рассмотрены ниже.

5. АНАЛОГОВЫЕ И АНАЛОГОВО-ЦИФРОВЫЕ НЕЙРОЧИПЫ

Некоторые характерные особенности ИНС, и прежде всего такие, как регулярность структуры и небольшое число выполняемых операций, дают идеальные предпосылки для построения нейрочипов в виде микросхемы. Как отмечалось выше, в настоящее время существуют три вида нейрочипов: аналоговые, аналогово-цифровые и цифровые нейрочипы. Каждый из них имеет свои плюсы и минусы, однако многое зависит от типа используемой модели ИНС. Здесь мы рассмотрим два характерных нейрочипа, которые подробно описаны в литературе, — аналоговый и аналогово-цифровой. Краткие сведения о других нейрочипах данного класса будут приведены ниже при рассмотрении их применения.

Аналоговые ИНС. По своей архитектуре аналоговые нейрочипы наиболее близки к естественным. Поэтому не случайно первые коммерческие разработки нейрочипов, получившие применение в научном приборостроении, предназначены для обработки аналоговых сигналов. Наиболее широкое применение в экспериментальной физике получил модуль фирмы «Intel» ETANN, описания которого приведены в [14] на русском языке и в [29] на английском языке (ETANN — Electrically Tranable Analog Neural Network — электрически обучаемая аналоговая нейронная сеть). Нейрочип ETANN (80170NX) выполнен в виде большой интегральной микросхемы, содержащей 208 выводов. Среди них 2×64 выводов предназначены для входных сигналов и 64 вывода

являются выходными, которые, в свою очередь, можно соединять с 64 входными контактами. Это позволяет с помощью внутреннего мультиплексора создавать рекуррентную сеть на 64 входа (выхода). Кроме того, возможны несколько вариантов построения каскадных сетей. С целью расширения функциональных возможностей сетей отдельные модули могут соединяться также параллельно или в виде дерева.

На кристалле микросхемы, выполненной по технологии электрически перепрограммируемых ячеек, сформировано 64 нейрона. В свою очередь, каждый нейрон имеет 64 информационных входа, 16 входов используются для задания величин начальных смещений. Схемы для задания смещений образуют матрицу 16×64 . Для хранения весовых коэффициентов имеется матрица синаптических элементов размерностью 80×64 . Всего 64 нейрона содержат 10240 синапсов. Каждая матрица имеет 80 входов, причем 16 входов используются для подачи напряжения смещения, а остальные входы являются информационными. Далее, каждый нейрон подключен к 16 фиксированным источникам смещения. Задержка сигналов в одном нейрочипе составляет 1–3 мкс. Величины весов запоминаются в электрически перепрограммируемых аналоговых ячейках, так что каждый вход модуля имеет свой вес, умножаемый на величину входного сигнала, который может меняться в пределах от 0 до 3 В с точностью 6–7 бит. Структура синапса нейрочипа обеспечивает умножение входного сигнала нейрона с учетом знака на хранящееся в памяти синапса значение весового коэффициента. Значения произведений независимо суммируются с помощью аналогового сумматора, выход которого соединен с входом устройства, моделирующего сигмоидную функцию, наклон которой можно программировать и контролировать. Эффективная вычислительная мощность модуля эквивалентна $4 \cdot 10^9$ операций с плавающей точкой в секунду, что сравнимо с мощностью современного суперкомпьютера. Важно также, что нейрочип ETANN наряду с аналоговыми сигналами при определенных условиях может использовать и цифровые сигналы для вычислений [14].

Аналогово-цифровой нейрочип ANNA. Для решения многих задач с помощью ИНС не требуется высокой точности. В то же время ряд функций ИНС, в том числе и программирование работы, более эффективно выполняется цифровыми методами. Поэтому весьма перспективными являются гибридные нейрочипы, в которых часть функций выполняется с помощью аналоговых устройств. Блок-схема и характеристики гибридного нейрочипа ANNA (Artificial Neural Network ALU — искусственный нейрочип с арифметико-логическим устройством) описаны в [30,31]. Модуль специально разработан для решения классификационных задач и прежде всего для распознавания рукописных текстов. Однако его архитектура позволяет реализовать также сеть Хопфилда. Модуль выполняет до $2 \cdot 10^6$ умножений и сложений в секунду и может производить до $5 \cdot 10^9$ соединений в секунду. Арифметические операции выполняются с точностью до 6 бит при вычислении весовых

коэффициентов, и 3 бита определяют состояния на выходах нейрочипа. Аппаратная система поддержки нейрочипа состоит из персонального компьютера и нейроплаты, выполненной в стандарте VME, на которой смонтированы нейрочип и цифровой сигнальный процессор. Цифровой сигнальный процессор используется для программирования нейрочипа и обучения сети. В нейрочипе содержится 4096 синапсов, которые с помощью мультиплексора позволяют создавать сети, имеющие сотни и тысячи соединений. С помощью команд можно запрограммировать количество синапсов в нейроне от 16 до 256. Тогда количество нейронов может составлять от 256 до 16. На рис. 9 приведена блок-схема нейрочипа. Входные 64-разрядные данные (по 3 бита на слово) в нейрочип заносятся последовательно по четырем независимым

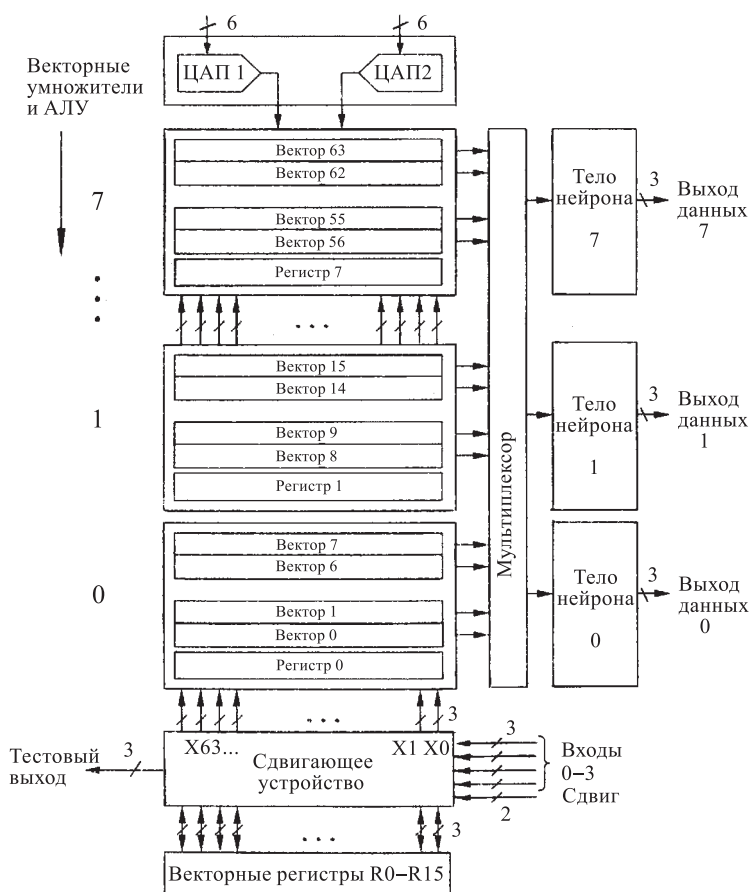


Рис. 9. Блок-схема нейрочипа ANNA

интерфейсам с частотой 120 Мбит/с. Основу нейрочипа составляют восемь векторных умножителей. В свою очередь, каждый умножитель содержит регистр, с помощью которого запоминается вектор состояния, и восемь АЛУ, каждое из которых подключено к 64 синапсам. С помощью мультиплексора осуществляются связи между выходами АЛУ и входами (телами) нейронов. Если содержимое регистров каждого АЛУ различно, то полученная сеть будет содержать максимальное количество нейронов — 256. Естественно, что архитектура нейрочипа позволяет программировать многослойные сети. Тело нейрона реализует сигмоидную функцию и преобразует результат операции в цифровое трехразрядное слово. Веса хранятся в виде электрических зарядов на конденсаторах. Для преобразования цифровых данных в аналоговые сигналы в процессе занесения весовых коэффициентов служат цифроаналоговые преобразователи (ЦАП1 и ЦАП2).

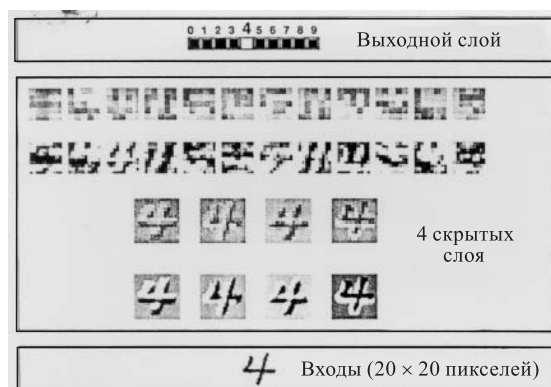


Рис. 10. Пример, иллюстрирующий состояния входного, четырех скрытых и выходного слоя сети для рукописной цифры 4 [30]. Сеть имеет 400 входов (20×20 пикселей). Состояния входов и четырех скрытых слоев имеют серый фон (цвет). Состояния на выходах соответствуют белому (правильный ответ) и черному цвету — на остальных выходах (отрицательный ответ)

Для экспериментов по распознаванию рукописных текстов была использована пятислойная прямооточная сеть, содержащая 20×20 входов и 10 выходов. На рис. 10 приведен результат эксперимента. Рисунок состоит из трех частей, характеризующих состояния входного, четырех скрытых и выходного слоя нейронной сети. Внизу показан входной образец (рукописная цифра 4). Средняя часть иллюстрирует процесс обработки информации в четырех скрытых слоях. В верхней части показано состояние выходного слоя. Нейрочип распознает до 20 символов в секунду. Для распознавания цифры 4 потребовалось 130 тысяч соединений.

6. НЕЙРОПЛАТЫ И НЕЙРОКРЕЙТЫ

Для эффективного использования нейрочипов необходимы системы тренировки и интерфейсы для цифровых процессоров общего применения. Поэтому наряду с нейромодулем ETANN фирма «Интел» выпускает аппаратную систему поддержки (систему моделирования) и гибридную вычислительную плату в стандарте VME [32]. Система позволяет тренировать одновременно до восьми модулей с помощью персонального компьютера i486. Разработан также гибридный процессор, смонтированный на двух платах. На одной плате установлена микроЭВМ 68070 с двухпортовой памятью емкостью 2 Мбайт, с двумя каналами прямого доступа и параллельно последовательный интерфейс для обмена данными с ПК. На другой плате расположены два модуля ETANN и два разъема для ввода аналоговых сигналов. Обмен данными между платами осуществляется с помощью прецизионных АЦП и ЦАП.

Цифроаналоговые преобразователи выполняют следующие функции: задают весовые коэффициенты, значения смещений и осуществляют программирование сигмоидной функции. С помощью АЦП выходы нейронной сети подключены к микроЭВМ. С учетом перспективности нейроплат в [33] проведено исследование каскадируемых гибридных плат, выполненных в стандарте VME. На рис. 11 приведена блок-схема экспериментальной системы, которая состоит из трех модулей ETANN, расположенных на плате VME. Такая структура получила название VME-3ETANN. Система может реализовать одно-, двух- или трехслойную ИНС. Такие конфигурации образуются путем использования стандартных интерфейсов J1, J2, расположенных на передней панели платы VME, и интерфейсов S1 и S2, находящихся непосредственно на плате. Например, используя только интерфейсы S1 и S2, можно построить трехслойную сеть. Разработанная архитектура позволяет создавать сети, имеющие кратное 64 количество входов и выходов. Поскольку модуль выполнен в стандарте VME, нет проблем со стыковкой аналоговых входов со вспо-

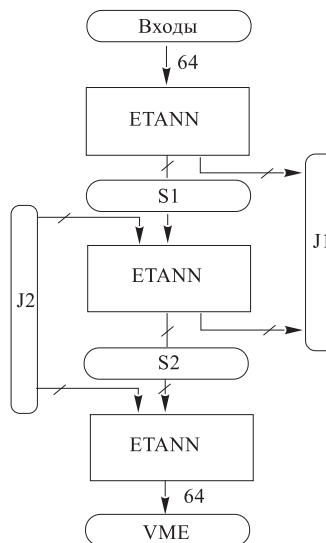


Рис. 11. Блок-схема стандартной нейроплаты VME-3ETANN: J1, J2, S1 и S2 — интерфейсы стандарта VME. Используя выходы разъема J2, можно реализовать одно- и двухслойную ИНС. Используя последовательно включенные модули ETANN с помощью интерфейсов S1 и S2, можно реализовать трехслойную ИНС. По остальным шинам, обозначенным кривой чертой, передаются 32-разрядные данные

могательными модулями, необходимыми для фильтрации входных данных и преобразования типа АЦП и ЦАП. Функция программы-драйвера сводится к выбору процессорной платы, установке моды работы (однонаправленная сеть или сеть с обратной связью), чтению или записи весов, установке адресов синапсов и т.д.

В [34] описан гибридный нейрокомпьютер, выполненный в виде крейта, содержащего 9 модулей в модернизированном стандарте VME [35] (версия VME9U с размерами платы 366,7×400 мм). Среди них два модуля содержат аналоговые нейрочипы и на двух модулях смонтированы цифровые нейрочипы типа M16 (см. ниже). Остальные модули содержат микропроцессоры, интерфейсы и контроллеры. Создание такой системы преследует три цели: использование нейрокомпьютеров для регистрации вторичных вершин распада в эксперименте WA92; исследование и сравнение возможностей аналоговой и цифровой ИНС в реальных экспериментах и внедрение гибридной нейросистемы в другие эксперименты. Для обеспечения наладки и работы гибридного нейрокомпьютера используется программа NEURAL [34], которая включает в себя подпрограммы обучения с помощью алгоритма обратного функционирования, подпрограмму вычисления коэффициентов для линейного дискриминатора Фишера и т.д.

7. ПРИМЕНЕНИЕ АНАЛОГОВЫХ И АНАЛОГОВО-ЦИФРОВЫХ НЕЙРОЧИПОВ В ЭКСПЕРИМЕНТАХ ПО ФИЗИКЕ ВЫСОКИХ ЭНЕРГИЙ

Обработка сигналов. С помощью ИНС эффективно решается проблема фильтрации сигналов на уровне значительных помех. Между цифровым фильтром и искусственным нейроном имеется прямая аналогия [29]. Принципиальным отличием искусственного нейрона от цифрового фильтра является наличие сигмоидной передаточной характеристики. Если к нейрону добавить элементы задержки и исключить нелинейность в передаточной характеристике, то получим формальный эквивалент фильтра. Это свойство нейрона позволяет использовать ИНС для обработки сигналов с применением быстрого преобразования Фурье или других методов. В [36] показано, каким образом нейрочип ETANN может быть использован в качестве фильтра сигналов. На рис. 12, *а* приведена трехслойная нейронная сеть [36]. Справа показаны два нейрона со смещением, а внизу — специальная аналоговая память последовательного типа с ответвителями. Количество нейронов в скрытом слое может меняться от 3 до 32. Показано, что чем больше нейронов в сети, тем «тоньше» настройка. Для настройки сети использовался алгоритм обратного функционирования (рис. 12, *б*), который в математической форме

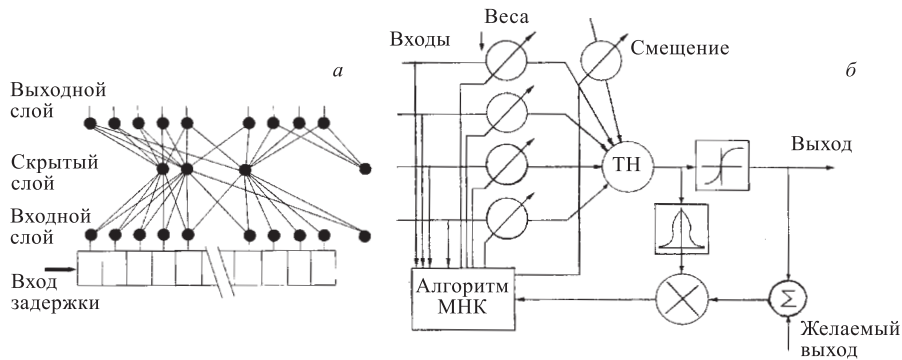


Рис. 12. Использование нейрочипа ETANN в качестве фильтра сигналов: *a*) блок-схема трехслойной однонаправленной нейронной сети (черными кружками обозначены нейроны); справа изображены элементы смещения; внизу — аналоговая линия задержки с ответвлениями; *б*) упрощенная схема обучения с помощью метода обратного функционирования для одного нейрона. ТН — тело нейрона; МНК — метод наименьших квадратов. Кривые изображают сигмоидную функцию и ее производную. Σ — сумматор, \times — схема умножения

(упрощенно) представлен в виде

$$f(\alpha) = 2/(1 + e^{-G\alpha}) - 1,$$

здесь G — параметр сигмоидной функции и

$$\alpha = \sum (a_i w_i + b_i w_i), \quad (2)$$

где a_i и b_i — i -й нейрон и i -й вход смещения соответственно и w_i — весовой коэффициент. Причем суммирование выполняется по всем связям, включая и элементы смещения b . Матрица весов получена с помощью алгоритма обратного функционирования. В данной работе использовался модифицированный вариант алгоритма обратного функционирования. Обычно для обучения используется пара векторов, состоящая из входного вектора и желаемого выходного вектора. Во время обучения эти данные подаются на входы сети, и веса меняются до тех пор, пока не получится желаемый результат. В предлагаемой версии алгоритма действительное значение входа используется также в качестве желаемого выхода. Далее, после первой итерации выбираются произвольные веса, и с помощью алгоритма обратного функционирования полу-

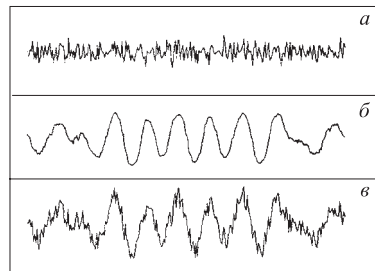


Рис. 13. Результат фильтрации спектра с использованием семи скрытых нейронов: *a*) вход фильтра в отсутствие сигналов; *б*) вид спектра после фильтрации; *в*) входные сигналы [36]

чаются новые веса для решения уравнения (2). С целью повышения точности настройки итерации повторяются. На рис. 13 показан результат фильтрации.

Примеры применения нейронной сети ETANN в экспериментах по физике высоких энергий. Необходимость в использовании ИНС в экспериментах по физике высоких энергий диктуется не только сложностью топологий редких событий, регистрируемых на уровне больших шумов, но и стремлением получить систему отбора частиц с высокой селективностью, так как полезное событие может содержаться среди $10^{10} - 10^{12}$ фоновых. Одним из преимуществ использования ИНС в экспериментах по физике высоких энергий является возможность распознавания сложных коррелированных событий между входными переменными, высокое быстродействие и высокая помехоустойчивость. Для решения задач, связанных с классификацией элементарных частиц, таких, как разделение легких и тяжелых кварковых струй, для идентификации частиц, а также в триггерных системах для выделения полезных событий на фоне помех, используются прямые ИНС [37].

Рассмотрим пример применения нейронной сети ETANN для обработки событий, регистрируемых в вершинном детекторе. Назначением вершинных детекторов в экспериментах по физике высоких энергий является нахождение и измерение времени жизни короткоживущих частиц порядка 10^{-10} с и менее. Топология таких событий имеет, как правило, одну или несколько вторичных вершин распада, которые пространственно отделены от первичной вершины и одна от другой на определенное расстояние, составляющее несколько миллиметров и менее. Объективным параметром, определяющим, ассоциируется ли заряженная частица со вторичной вершиной, является прицельный параметр P (или «промашка»), поскольку его значение не зависит от импульса частиц и поэтому не требуются дополнительные измерения. Так, если частица исходит прямо из точки первичной вершины распада, то «промашка» $P = 0$. В [38] гибридный процессор с нейрокомпьютером ETANN используется для детектирования вторичной вершины распада частиц в эксперименте с фиксированной мишенью по поиску тяжелых цветных кварков. Исследование проводилось при условии, что отношение сигнал/шум равно 1:50000. На рис. 14 приведена схема установки. Частицы, исходящие из мишени и вторичной вершины распада, детектируются с помощью микрополоскового полупроводникового телескопа. После усиления и оцифровки сигналов координаты частиц поступают на входы ассоциативных запоминающих устройств, с помощью которых отбираются только заведомо полезные треки, имеющие заданный наклон. Прицельный параметр определялся путем использования предварительно обученной нейронной сети. В табл. 1 приведены сравнительные данные, полученные с помощью моделирующей нейронной сети программы и с помощью гибридного процессора.

Как следует из табл. 1, процент полезных событий, идентифицированных с помощью нейронной сети ETANN, несколько хуже, нежели это делает программа

Таблица 1. Процент полезных, фоновых событий и шумов, которые идентифицированы верно, неверно или неопределенно с помощью моделирующей нейронную сеть программы и аппаратной сети, созданной на основе нейрочипа ETANN

Программа	Полезные соб.	Фоновые соб.	Неопределенные соб.
Сигнал	57 %	37 %	6 %
Шум	92 %	6 %	2 %
ETANN			
Сигнал	48 %	48 %	4 %
Шум	92 %	6 %	2 %

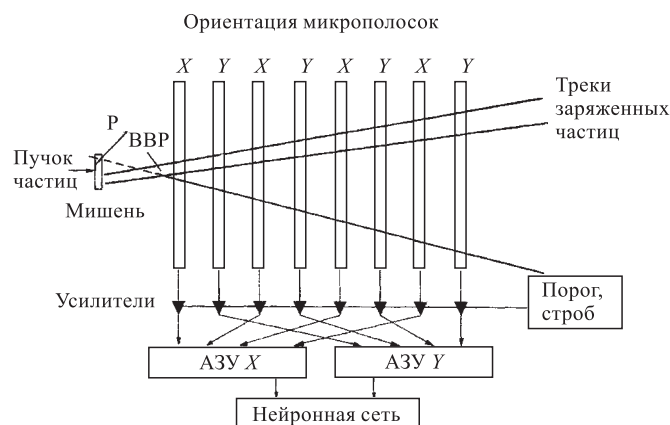


Рис. 14. Блок-схема установки с нейронной сетью для регистрации координат вершины распада вторичных частиц. ВВР — вторичная вершина распада; АЗУ — ассоциативное запоминающее устройство. Если продлить трек, исходящий из ВВР, в сторону мишени, то он пересечет ее в определенной точке плоскости мишени. Перпендикуляр, опущенный из этой точки на ось X , будет равен прицельному параметру P . С помощью ассоциативных запоминающих устройств регистрируются треки с заданным наклоном. Функция обученной ИНС сводится к регистрации только тех событий, которые имеют в пределах заданной точности прицельный параметр, отличный от нуля

при одинаковых значениях процентов фильтрации шумов. В то же время с помощью процессора гораздо лучше фильтруются фоновые события, т.е. такие события, которые не содержат вторичной вершины распада при одинаковых значениях фильтрации шумов. Процент распознавания неопределенных событий практически одинаков для обоих методов. В [39] описано использование ИНС, выполненной на основе нейрочипа ETANN в качестве фильтра сигналов, получаемых с помощью жидкоаргонной времяпроекционной камеры.

ИНС тренировалась путем подачи спектра сигналов, регистрируемых в сегменте, содержащем 192×2048 каналов. Функции фильтра заключались в том, чтобы обеспечить высокое отношение сигнал/шум в процессе определения позиции пика и восстановления треков заряженных частиц. Настройка сети выполнялась в три этапа. Вначале для обучения моделировались изображения треков на уровне шумов. После этого проверялась работа программы фильтрации. После получения удовлетворительного результата составлялись таблицы весов, параметров сигмоидной функции и величин смещений. Затем эти данные использовались для обучения аппаратной ИНС. На рис. 15 приведены результаты фильтрации. В верхней части рисунка показана картинка до фильтрации. В нижней части рисунка отчетливо видны треки частиц, которые стали заметными после фильтрации. Применение ИНС позволило повысить скорость обработки сигналов в 10^6 раз по сравнению с обычными методами реконструкции треков заряженных частиц.

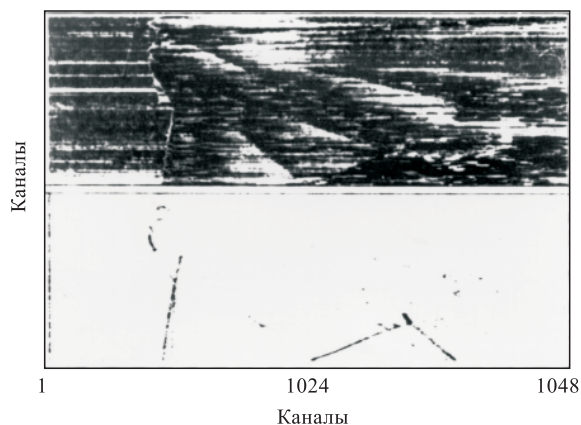


Рис. 15. Пример фильтрации треков заряженных элементарных частиц, зарегистрированных с помощью жидкоаргонной времяпроекционной камеры: верхняя часть — картинка до фильтрации, которая была получена с помощью математического фильтра, внизу — после фильтрации. Данные были получены от фрагмента детектора, содержащего спектр для 192×2048 каналов [39]

Возможность использования нейронной сети ETANN для распознавания событий, регистрируемых с помощью сцинтилляционного калориметра в Лаборатории им. Э.Ферми, была показана в [40]. Затем на базе данного нейронной сети была создана система выработки импульса запуска установки (триггерная система), где в качестве детектора используется электромагнитный калориметр [41]. Функция триггерной системы заключается в том, чтобы с помощью зарегистрированного изолированного кластера идентифицировать: а) наличие изолированного фотона в центральной области калориметра; б) факт

регистрации изолированного электрона или фотона в торце калориметра; в) обнаружение полупетонного b -распада. Нейрочип ETANN был выбран с целью реализации алгоритма регистрации изолированной частицы в триггерной системе второго уровня. Калориметр разделен на 25 электромагнитных и 25 адронных частей. Причем каждой части соответствует один входной нейрон. Сигналы на каждый скрытый нейрон поступают от выходов пятидесяти нейронов входного слоя (рис. 16). Эффективность регистрации составляет 97%. На рис. 17 приведена картинка с дисплея, на которой изображено полезное событие, которое не регистрировалось с помощью ранее используемого обычного алгоритма без применения нейрочипа. Черные фигуры обозначают теряемую электронами энергию в торцевой части калориметра.

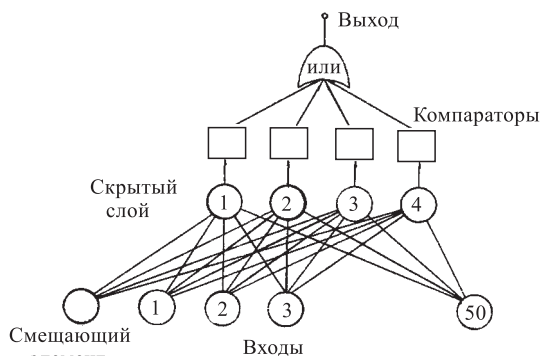


Рис. 16. Блок-схема части однонаправленной ИНС для вычисления алгоритма регистрации изолированной частицы с помощью сцинтилляционного калориметра. Сигналы на входы сети подаются от 25 электромагнитных и 25 адронных сегментов калориметра

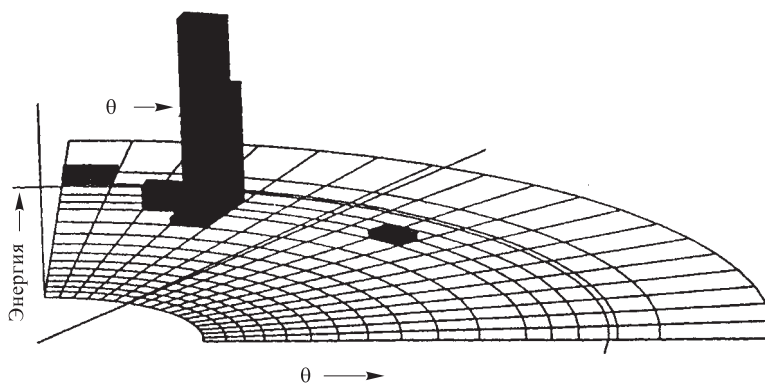


Рис. 17. Пример события, зарегистрированного с помощью ИНС. Черные фигуры изображают энергию в продольном направлении, теряемую электронами в торцевой части калориметра. Максимальное значение энергии равно 14,4923 ГэВ [41]

Известен ряд работ, посвященных использованию аналогового нейрочипа ETANN для идентификации элементарных частиц. В [42] нейрочип входит в

состав устройства, предназначенного для выработки импульса, запускающего установку, с целью регистрации и идентификации тау-лептонов. Данные на входы сети подаются от сегмента калориметра, содержащего $5 \times 5 \times 2$ ячеек. Сеть имеет структуру 50-50-1. В [43] в качестве детектора используется черенковский детектор кольцевого изображения (RICH-детектор) с целью отделения протонов от пионов. Основу экспериментальной установки составляет черенковский радиатор из кристаллов NaF толщиной 10 мм и оригинальный светочувствительный детектор. Световой конус, эмитируемый из радиатора, отражаясь от кристаллов, экспонируется в виде кольца черенковского света в дрейфовой области, наполненной гелием. Далее черенковские фотоны проходят сквозь окно в кварцевом стекле и конвертируются в электроны в светочувствительном газе. Затем фотоэлектроны дрейфуют в направлении чувствительных проволочек многопроволочной пропорциональной камеры. Как следствие этого, на электродах многопроволочной пропорциональной камеры (МПК) индуцируются электрические сигналы. Катодная плоскость МПК состоит из 12×12 микрополосок (8×8 мм каждая) квадратной формы, что позволяет получать две координаты для каждого фотоэлектрона. Обычно от одной частицы срабатывают 2–5 микрополосок, поэтому использовались данные, поступающие от 128 датчиков, что позволило задействовать один модуль типа ETANN. При импульсе 3,5 ГэВ/с скорости протонов и пионов равны, соответственно, 0,966с и 0,992с (с — скорость света в вакууме).

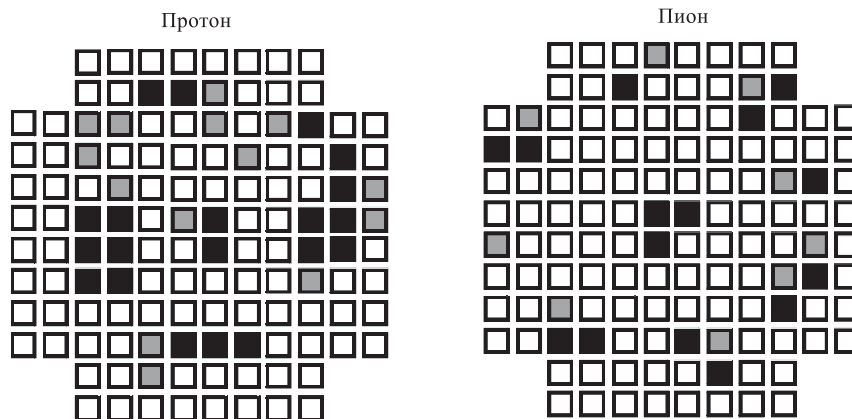


Рис. 18. Образцы событий, используемых для тренировки ИНС. 128 аналоговых сигналов одновременно подаются на входы ИНС. Затемнения показывают логарифмическое изменение интенсивности. В ИНС это соответствует напряжению в пределах 0–3 В [43]

В то время как более медленные частицы (протоны) дают плотные и меньших размеров кольцевые изображения, пионы генерируют аналогичные

изображения больших размеров с тонкими и прерывистыми линиями. Типичный размер кольца составляет 38 мм. На рис. 18 приведены образцы, используемые для настройки сети, причем приведены модели событий с протонами и пионами в таком виде, в каком они регистрируются с помощью стрипов детектора. Штриховка отображает изменение интенсивности в логарифмическом масштабе. При этом используются пять градаций интенсивности, что соответствует изменению входного сигнала от 0 до 3 В.

На рис. 19 приведено схематическое изображение нейронной сети, включающее предусилители и логарифмические усилители. Напряжение смещения подается на входы скрытого и выходного слоя. Если детектируется протон, то на выходе сети «Выход(p)» выходного слоя устанавливается высокий уровень напряжения, а на выходе «Выход(π)» будет низкий уровень напряжения. Для формирования сигмоидной функции используется алгоритм обратного функционирования. Показано, что применение гибридного процессора, состоящего из нейрочипа ETANN и микроЭВМ 68070, позволило существенно повысить достоверность распознавания событий от 87 %, когда использовалась математическая модель нейронной сети, до 98 % в случае применения аппаратной нейронной сети.

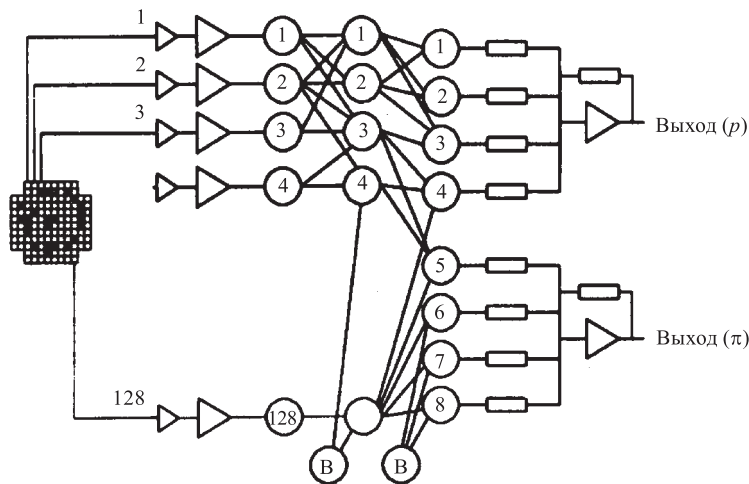


Рис. 19. Блок-схема аналоговой ИНС, используемой для идентификации протонов и пионов с помощью детектора кольцевого изображения. «Выход(p)» и «Выход(π)» — выходы сети; 1÷128 — входы сети; В — элементы смещения; остальные кружки обозначают нейроны. Слева на рисунке изображен образец, подаваемый на входы сети [43]

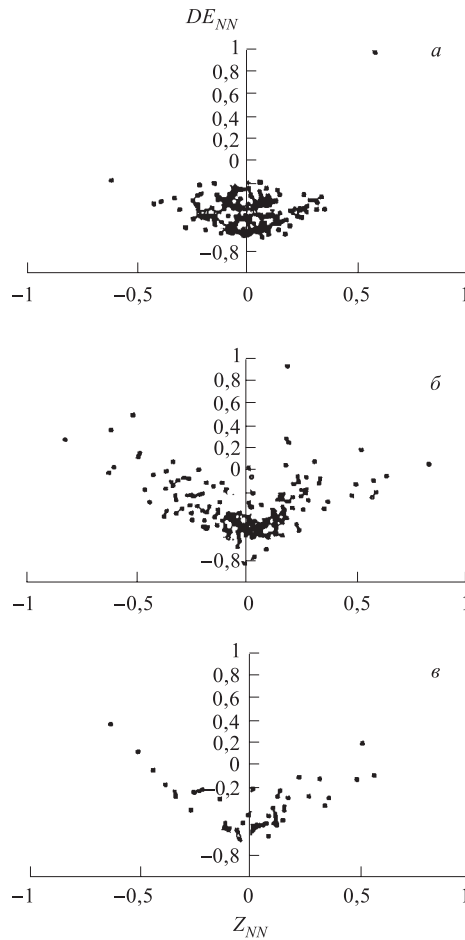


Рис. 20. Распределения входных значений DE_{NN} и Z_{NN} , распознаваемых с помощью ИНС: для каонов (а) и для пионов (б); распределение для некорректной классификации каонов и пионов (в). DE_{NN} — сумма сигналов, формируемых на концах сцинтилляторов, обеспечивающих в оцифрованном виде информацию о потере энергии частицами в сцинтилляторах; Z_{NN} — разница этих сигналов, которые дают информацию о z -координате зарегистрированной частицы

С целью идентификации каонов и пионов в [44] используется четырехслойная ИНС с двумя скрытыми слоями и смещением. В качестве детектора служат сцинтилляционные счетчики большой длины. Сигналы, поступающие с выходов ФЭУ, оцифровываются с помощью параллельных АЦП. Далее формируются два типа сигналов: DE_{NN} — сумма сигналов, формируемых на концах сцинтилляторов, и Z_{NN} — разница аналогичных сигналов. Эти данные подаются непосредственно на входы сети, выполненной на основе ставшей уже стандартной платы VME-3ETANN.

Идентификация рассматриваемых частиц возможна, поскольку каоны и пионы имеют различные величины потерь энергии dE/dx на единицу длины. Для выбора окончательной архитектуры ИНС с помощью алгоритма обратного функционирования было использовано 2000 тестовых векторов, соответствующих идентифицируемым частицам. На рис. 20 приведены результаты идентификации с помощью обычной сети, в которых использовалось по 450 векторов, соответствующих данным о каонах и пионах. На рис. 20, а и рис. 20, б приведены распределения, соответствующие каонам и пионам. С помощью рис. 20, в иллюстрируется отрицательный результат, поскольку идентификация частиц невозможна.

Перспективы использования ИНС в экспериментах на большом адронном коллайдере. Поиск гипотетической частицы — хиггс-бозона начнется,

скорее всего, не ранее 2005 года, после успешного запуска ускорителя. Поэтому с целью подготовки к анализу информации, регистрируемой с помощью сложной системы детекторов, и с учетом редкости событий в [45] проведено исследование по использованию ИНС для поиска хиггс-бозона. В качестве образцов использовались коды, генерируемые методом Монте-Карло. Требовалось также сравнить, насколько аппаратные методы более эффективны по сравнению с программными ИНС, и оценить возможности двух типов аналоговых чипов для решения поставленной задачи. В качестве нейрочипов использовались аналоговые модули ETANN, CLN32 и вспомогательный чип CLN64 [45].

Рассмотрим кратко особенности аналогового нейрочипа CLN32. В отличие от многих других нейрочипов на кристалле содержится встроенное устройство для обучения с помощью двух независимых алгоритмов, один из которых реализует машину Больцмана. В отличие от нейрочипа ETANN, в модуле CLN32 веса запоминаются в цифровой форме. Имеется также 32 нейрона и 496 адаптивных синапсов. Кроме того, на отдельном чипе CLN64 имеется 1024 адаптивных синапса, что позволяет существенно расширять функциональные возможности основного нейрочипа. Задержка сигналов в однослойной сети составляет всего 1 мкс. В процессе обучения данные на 32 аналоговых входа подавались с частотой 50 кГц. Для исследований была выбрана прямочная сеть с архитектурой 8-16-2 для программного эмулятора и для нейрочипа ETANN. Это значит, что входной слой имеет восемь нейронов, шестнадцать нейронов содержится в скрытом слое и два нейрона имеется на выходах. Необходимость в двух двоичных выходах с четырьмя различными состояниями диктуется тем, что для идентификации событий в данном случае требуется генерация трех сигналов на выходах: «сигнал события», «событие, не представляющее интерес» и «фоновое событие». На основе нейрочипа CLN32 исследовалась топология ИНС 24-16-2. Исследования проводились по следующей программе: возможности программной эмуляции и сравнение двух типов нейрочипов. В результате исследований показано, что путем применения алгоритма обратного функционирования качество отбора событий с помощью программного эмулятора и с помощью нейрочипа ETANN практически одинаково. Величина задержки ИНС составляет 8 мкс, что не позволяет использовать такую сеть в быстрых устройствах для отбора частиц. Кроме того, по сравнению с нейрочипом ETANN модуль CLN32 допускает гораздо большее число ошибок в процессе классификации частиц.

В связи с этим представляет интерес аналогово-цифровой нейрочип MESA [46], время установки данных в котором не превышает 50 нс. Архитектура нейрочипа позволяет создавать прямочную сеть, имеющую 70 входных и 6 скрытых нейронов и несколько выходов. Последнее качество достигается тем, что взвешенные входы каждого нейрона скрытого слоя имеют внешние выводы. В то время как входные сигналы являются аналоговыми, пятираз-

рядные синаптические веса программируются обычным способом с помощью внешнего процессора.

Нейрочип MESA используется в триггерной системе первого уровня с целью отбора событий, по своим характеристикам весьма близких к фоновым в эксперименте Н1, проводимом на ускорителе со встречными пучками HERA (Гамбург). Частота поступления фоновых событий составляет 100 кГц при частоте регистрации полезных событий порядка 10 Гц. Требуется отбор событий, имеющих первичную вершину распада в точке встречи пучков протонов с энергией 820 ГэВ и позитронов с энергией 27,5 ГэВ. Причем ступки частиц сталкиваются каждые 96 нс. Для построения триггерной системы использовались два нейрочипа, поскольку гистограмма событий имеет 96 входов. Для тестирования ИНС на линии с персональным компьютером была выбрана VME-плата. С целью максимального приближения к реальным условиям наряду с полезными сигналами в систему вводились шумы. Помимо нейрочипа на тестируемой плате имеется 70 ЦАП для преобразования входных цифровых сигналов в аналоговые и 12-разрядный АЦП, используемый для чтения выходных сигналов. Показано, что классификации полезных и фоновых событий с помощью ИНС достаточно для их разделения. Причем эффективность сети повышается с ростом числа входов.

8. ПРИМЕНЕНИЕ АНАЛОГОВЫХ ИНС ДЛЯ ИЗМЕРЕНИЯ ФИЗИЧЕСКИХ ПАРАМЕТРОВ И УПРАВЛЕНИЯ ОБЪЕКТАМИ

Распознавание образов и классификация объектов являются типичными задачами, решаемыми с помощью ИНС. При решении задач распознавания образов с помощью ИНС возникает проблема ввода желаемых образцов во входной слой ИНС. Однако в экспериментальной технике имеется ряд задач, когда такая проблема не возникает. В [47] была предложена идея использования аналоговых ИНС для измерения физических параметров. Такими выходными параметрами могут быть, например, угол наклона трека заряженной частицы или положение пучка заряженных частиц относительно оси и его ширина на полувысоте. В качестве входных данных могут быть координаты событий или временные параметры. При этом сигналы можно подавать непосредственно на входы ИНС или с помощью усилителей.

Для реализации рассмотренной идеи возможны два метода (подхода). Реализация первого метода предполагает использование в выходном слое сети только одного нейрона, выходное значение которого разделено на несколько интервалов. К недостатку такого подхода можно отнести то, что если нейрон имеет нелинейную характеристику, то в этом случае трудно оценивать чувствительность ИНС как прибора. Отрицательно сказывается также нестабильность значений напряжений на выходе. Суть альтернативного метода

заключается в том, что в выходном слое используется как можно больше нейронов с соответствующими выходами. В этом случае каждому из выходов сети можно сопоставить определенный, присущий ему интервал измеряемой величины. Для этой цели нейрочип ETANN является подходящим модулем, поскольку он имеет 64 выхода. В данной работе обсуждается вопрос выбора параметра x в выражении для сигмоидной функции (см. рис. 3), который обеспечивал бы работу ИНС в качестве измерительного прибора в заданных пределах.

В [48] описана экспериментальная установка, предназначенная для измерения параметров треков заряженных частиц, регистрируемых с помощью газонаполненной дрейфовой камеры. Описание используемого детектора читатель может найти в цитируемой работе. В качестве входных сигналов сети используются время дрейфа электронов к сигнальной проволочке (одна координата частицы) и позиции микрополосок в катодной плоскости (другая координата частицы). Трехслойная ИНС, созданная на основе нейрочипа ETANN, имеет структуру 64-64-64. Причем входы и выходы ИНС разделены на две равные группы. На первую группу входов подаются сигналы от времязадающих генераторов, и на вторую поступают усиленные сигналы от микрополосок детектора. Первой группе выходов соответствует расстояние в детекторе, равное 20 см, так что один бин равен 0,625 см. В свою очередь, второй группе выходов соответствует угол наклона трека заряженной частицы в пределах 45° . На основе этих данных ИНС тренировалась по методу обратного функционирования. Получено разрешение координаты трека 0,9 мм и угла наклона 16 мрад, в то время как при моделировании на ЭВМ получены худшие данные: 1,2 мм и 20 мрад соответственно.

В [49] ИНС на основе нейрочипа ETANN используется для определения положения пучка частиц и его ширины на полувысоте. В отличие от предыдущей работы сеть имеет всего два аналоговых выхода, которые отградуированы соответствующим образом.

С помощью аналоговых и аналогово-цифровых ИНС эффективно решается ряд сложных проблем, связанных с нелинейным управлением. В [50] разработанная авторами ИНС используется в токамаке для контроля формы и положения плазмы в эксперименте COMPASS. На рис. 21 приведена блок-схема системы контроля. Сигналы на входы ИНС поступают от электромагнитных датчиков, окружающих вакуумную камеру. В процессе обработки этих данных на выходах нейрокомпьютера получают измеренные параметры плазмы y_k , которые описывают координаты и форму плазмы на границе вакуумной камеры. Путем сравнения этих величин с заданными получают сигналы погрешности, которые по цепям обратной связи поступают в корректирующие обмотки управляющих устройств. Высокий параллелизм вычислений, присущий аналогово-цифровой ИНС, позволяет быстро, за несколько десятков микросекунд, посылать корректирующие сигналы. Синаптические

веса в аналоговом виде получаются с помощью ЦАП, которые, в свою очередь, программируются.

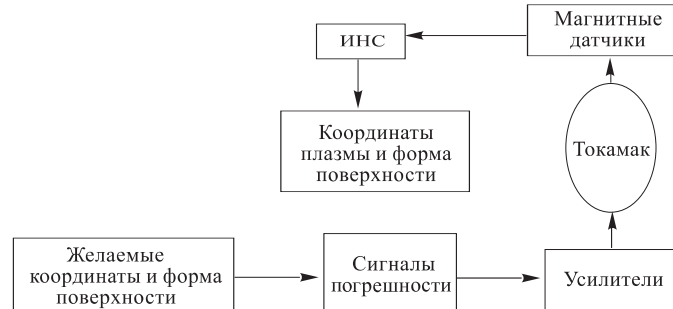


Рис. 21. Блок-схема контроля позиции и формы плазмы в токамаке. ИНС осуществляет нелинейное преобразование измеряемых сигналов магнитных датчиков в множество геометрических параметров, которые описывают координаты и форму плазмы. Затем эти параметры сравниваются с желаемыми параметрами и результат погрешности используется для коррекции тока в корректирующих обмотках

Другим примером применения ИНС в системах управления может служить установка для контроля положения шести сегментов зеркала астрономического оптического телескопа [51]. В адаптивной оптике телескопа примерно каждые 10 мс требуется изменение эффективной поверхности зеркала с целью подавления атмосферных помех. Обычный метод решения этой задачи заключается в использовании итеративного алгоритма, для реализации которого требуются трудоемкие вычисления с целью коррекции поверхности зеркала. Эксперимент по измерению искажения оптической фазы, возникающей вследствие турбулентности атмосферы, проводился в реальных условиях с использованием света от звезды. Использование ИНС (название ИНС, к сожалению, авторы не указали) обеспечивает быстрое и с необходимой точностью управление формой зеркала, что позволило повысить качество изображения объекта наблюдения. Данные на входы ИНС подавались от камеры с ПЗС-структурой. В [52] представлен обзор применения ИНС для решения различных задач управления, написанный по материалам зарубежной печати последних лет.

9. ПРИМЕНЕНИЕ ТРАНСПЬЮТЕРОВ И ЦИФРОВЫХ СИГНАЛЬНЫХ ПРОЦЕССОРОВ

В настоящее время сложились два направления, характеризующие развитие нейроконтроллеров: разработка ускорительных плат на базе БИС и

микропроцессоров и реализация нейрокомпьютеров на базе универсальных многопроцессорных систем. Первое направление получило наиболее широкое применение в научном приборостроении, и поэтому мы остановимся на нем более подробно.

Среди аппаратной поддержки нейровычислений следует отметить применение быстрых микропроцессоров с RISK-архитектурой (например, транспьютеры), цифровых сигнальных процессоров (ЦСП) и специальных нейропроцессоров и вычислителей на их основе.

Применение транспьютеров. Ряд характерных особенностей транспьютеров обусловили их широкое применение для построения мощных и относительно дешевых вычислительных сетей, в том числе и для построения нейрокомпьютеров. В типичном модуле на одном кристалле (рис. 22) содержится высокопроизводительный микропроцессор с плавающей точкой и второй процессор с RISK-архитектурой, обуславливающий эффективность транспьютера. Кроме того, транспьютер содержит быструю оперативную память; и, что весьма важно, наряду с параллельным каналом прямого доступа в память имеются четыре последовательных дуплексных канала обмена (линки), с помощью которых можно осуществлять гибкий обмен информацией с соседними транспьютерами, как по горизонтали, так и по вертикали со скоростью 20 Мбит/с и более. С помощью линков можно создавать различные по топологии и производительности сети типа матричной, «дерево», конвейерные многопроцессорные системы и т.д. Важно, что каждый транспьютер в сети может работать в режиме ведущего и ведомого. В научном приборостроении транспьютеры широко используются для построения ускорительных плат, выполняемых, как правило, в стандарте VME. Так, ускорительная плата [53] содержит два транспьютера T800 и память емкостью 4 Мбайт.



Рис. 22. Блок-схема типичного транспьютера. В одном кристалле содержится процессор с плавающей точкой и второй процессор с RISK-архитектурой, оперативную память. Наряду с каналом прямого доступа к памяти имеется четыре последовательных дуплексных канала обмена (ПИ — последовательный интерфейс)

Важно, что каждый транспьютер в сети может работать в режиме ведущего и ведомого. В научном приборостроении транспьютеры широко используются для построения ускорительных плат, выполняемых, как правило, в стандарте VME. Так, ускорительная плата [53] содержит два транспьютера T800 и память емкостью 4 Мбайт.

Одна из последних версий фирмы INMOS — транспьютер T9000 содержит еще дополнительный коммуникационный процессор, КЭШ-память, пятиступенчатый конвейер для выполнения команд, мультиплексный канал и т.д., позволяет выполнять 25 млн операций с плавающей точкой в секунду [54]. В данной работе описаны также нейрокомпьютеры, созданные на базе транспьютеров серии T9000. Транспьютеры используются для управления каскадируемой нейровычислительной системой на базе нейрочипа L-Neuro, разработанного фирмой «Telmat» (Франция) [55].

Нейронная сеть на базе транспьютера T800 [56] использовалась для идентификации электромагнитных и адронных ливней, регистрируемых с помощью калориметра. В качестве модели использовалась прямоочная нейронная сеть, содержащая четыре входных нейрона, три нейрона содержатся в скрытом слое и один нейрон имеется в выходном слое. В состав аппаратуры входят миникомпьютер VAX3000, персональный компьютер и два транспьютера T800. Персональный компьютер в основном выполняет функции сервера и устанавливает значения весов в транспьютерах. Эффективность дискриминации параметров частиц составила не менее 96%.

Применение цифровых сигнальных процессоров. Цифровой сигнальный процессор типа TMS320C40 широко используется в экспериментах по физике высоких энергий вследствие высокой производительности и гибкости архитектуры. Он выполняет до 40 млн. операций с плавающей точкой в секунду, имеет два 32-разрядных интерфейса и шесть 8-разрядных интерфейсов с прямым доступом к оперативной памяти, с помощью которых данные передаются со скоростью 20 Мбайт/с. Наличие такого количества интерфейсов позволяет создавать не только многомерные топологические структуры типа гиперкуба, но и использовать его в качестве коммутатора данных в системах запуска установок, предназначенных для экспериментов на большом адронном коллайдере [57]. Кроме того, ЦСП TMS320C40 имеет следующие параметры: 32-разрядная сетка, 12 регистров для временного хранения данных, два модуля двухпортовых ОЗУ емкостью $1\text{К} \times 32$. Процессор может выполнять в течение одного цикла операцию умножения с плавающей точкой. Наличие четырех внутренних магистралей позволяет с помощью одной команды в течение одного цикла выполнять такое часто встречающееся выражение, как $Z = Y + AX$. Такие выражения, как $1/X$ и $1/X^{1/2}$, выполняются табличным способом [58].

Поэтому не случайно на базе процессоров TMS320C40 разработаны высокопроизводительные наращиваемые вычислительные модули типа M1 и M2, предназначенные для обработки информации в реальном времени, в том числе видеоизображений, радиолокационной информации, сигнальной обработки и т.д. Модуль M1 со стандартным интерфейсом ISA, который используется в микропроцессорах (Industrial System Architecture), содержит четыре процессора с тактовой частотой 50 МГц и имеет производительность

200 Мфлопс с объемом статической памяти 5 Мбайт. Модуль *M2* имеет быстродействие до 300 Мфлопс, представляет собой плату с интерфейсом VME и содержит шесть процессоров. Для повышения производительности можно наращивать несколько модулей [59]. Ниже будет показано, каким образом ЦСП TMS320C40 используется для построения ускорительных нейронных плат.

10. ЦИФРОВЫЕ НЕЙРОЧИПЫ И НЕЙРОКОМПЬЮТЕРЫ

В [60] описан полностью цифровой программируемый нейрочип (название нейрочипа в статье не приводится) Научно-технического центра (НТЦ) «Модуль», имеющий ряд рекордных параметров и оригинальную архитектуру.

При создании архитектуры нейрочипа разработчики стремились обеспечить программируемость узлов, специально разработанных для поддержки нейровычислений: параллельность вычислений и обеспечение вычислений над данными (весами) произвольной разрядности (от 1 до 64 бит). Нейрочип предназначен для реализации нейросетей. В качестве примера рассмотрим принцип действия операционного узла нейрочипа, предназначенного для вычисления весовых коэффициентов.

Эти операции требуют выполнения множества часто повторяющихся операций умножения и сложения в процессе работы нейрочипа. Особенностью функционирования операционного узла является то, что всего лишь за один такт осуществляется вычисление весовых коэффициентов нескольких синапсов и вычисление их сумм для нескольких нейронов. Работа операционного узла иллюстрируется на примере реализации функции фрагмента слоя нейросети, включающего два нейрона и три синапса (рис. 23), где $m = 19$ — общее число разрядов выходов, $n = 12$ — суммарная разрядность входов. Операнды имеют следующие разрядности: $X_1 = 3$, $X_2 = 4$, $X_3 = 5$; $W_{11} = 5$, $W_{12} = 4$, $W_{13} = 3$, $W_{21} = 4$, $W_{22} = 3$, $W_{23} = 2$; $U_1 = 10$, $U_2 = 9$, где U_1 и U_2 — смещения; $Y_1 = 10$ и $Y_2 = 9$.

Вначале в операционный узел загружаются элементы матрицы весов, содержимое которой вычисляется в процессе обучения нейросети. Элементами матрицы являются разряды весовых коэффициентов и нулевые элементы,

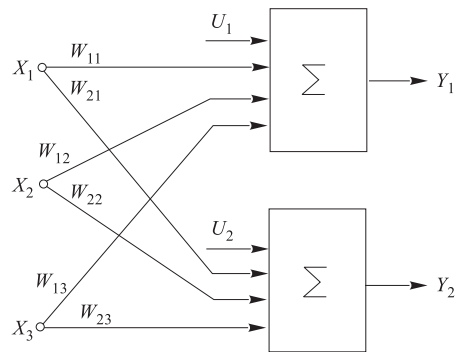


Рис. 23. Иллюстрация работы операционного узла (ОУ) нейрочипа на примере фрагмента слоя нейрона, состоящего из двух нейронов и трех синапсов. $X_1 \div X_3$ — входы; U_1, U_2 — смещение, $W_{11}, W_{21}, W_{12}, W_{22}, W_{13}, W_{23}$ — весовые коэффициенты; Y_1, Y_2 — выходы

расположенные в строго фиксированных позициях. Такая настройка как бы разбивает операционный узел на подматрицы размером $P \times R$ (рис.24), где P — количество нейронов, R — количество синапсов, во фрагменте сети. Для рассматриваемого примера $P = 2$, $R = 3$. Пусть подматрица

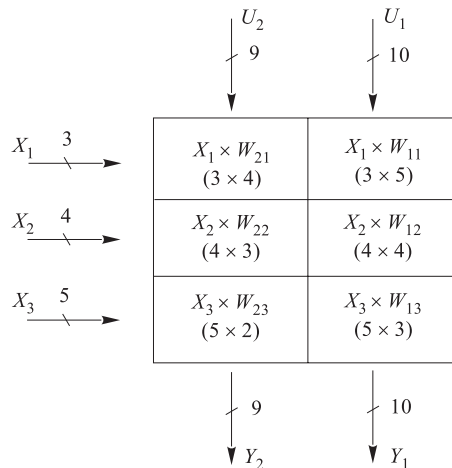


Рис. 24. Структура подматрицы. Все обозначения аналогичны обозначениям рис. 23 (см. текст)

с координатами (p, r) соответствует p -му, $p = 1, \dots, P$, нейрону и r -му, $r = 1, \dots, R$, синапсу. В результате получается, что каждая подматрица с координатами (p, l) вычисляет функцию $X_l W_{pl} + U_p$, а каждая подматрица с координатами (p, t) реализует произведение $X_l W_{pt}$ и прибавляет к нему результат, полученный на выходе подматрицы с координатами $(p, t - 1)$, где $t = 2, \dots, R$. В результате на выходе подматрицы (p, R) получается

$$Y_p = U_p + \sum X_r W_{pr}.$$

Следует отметить, что количество нейронов, обрабатываемых одновременно, зависит от числа входов, их разрядности и значений весовых коэффициентов. При работе нейрочипов с одинаковыми 8-разрядными данными операционный узел за один процессорный такт выполняет 24 операции умножения с накоплением. При частоте тактовых импульсов 30 МГц это составляет 720 млн операций с плавающей точкой в секунду (MCPS). В цитируемой работе приводится блок-схема нейрочипа и ее описание. Для настройки нейрочипа с конкретными величинами имеются специальные регистры. Существенными узлами нейрочипа являются также скалярный и векторный процессор. Наличие двух двунаправленных портов, аппаратно совместимых с портами ЦСП TMS320C40, позволяют создавать различные высокопроизводительные системы: двунаправленный конвейер, двухмерная и трехмерная решетка, двунаправленное кольцо, дерево и т.д.

Нейрокомпьютер SYNAPS1 [61]. Данный нейрокомпьютер разработан фирмой «Siemens» и предназначен для решения широкого круга задач, таких, как распознавание символов и обработка более сложных образов, например, шахматных фигур, речевых сообщений, управление роботами и т.д. Основу нейрокомпьютера составляет матричный процессор, состоящий из восьми сигнальных процессоров, выполняющих 3,2 млрд операций умножения (16×16 бит) и сложения. Емкость памяти составляет 32 Мбайт, скорость

передачи данных 0,8 Гбайт/с по шине данных, 200 Мбайт/с по управляющей шине и 56 Мбайт/с по адресной шине с тактовой частотой 25 МГц. Кроме того, нейрокомпьютер включает память весов емкостью 128 Мбайт и два управляющих микропроцессора MC 68040 фирмы «Моторола».

Нейрочип MA16. Особенностью нейрочипа является то, что его архитектура позволяет организовать систолическую вычислительную архитектуру. Данный тип процессорных структур получил название систолических, исходя из той аналогии, которая существует между операцией в базовом процессоре и сокращением сердца, а также между пульсациями потоков исходных данных, промежуточных и окончательных результатов и кровотоков [62]. Хотя систолические процессоры представляют собой новый класс конвейерных архитектур, имеется то принципиальное отличие, что в систолической структуре имеются три признака: конвейерность, синхронность и регулярность. Как следствие этого, отдельные процессоры работают параллельно, выполняя базовую операцию, после чего осуществляется синхронная передача выходных данных от одного процессорного элемента к соседнему по всем локальным связям. При этом данные в систолическом процессоре распространяются по конвейеру, обеспечивая многократное использование промежуточных данных, поступающих на вход процессора.

Нейрочип MA16 используется в эксперименте WA92 по поиску прелестных частиц совместно с аналоговым нейрочипом ETANN (см. выше) [34]. Нейрочип MA16 представляет собой цифровой систолический микропроцессор, разработанный фирмой «Siemens» [34,63]. Его архитектура позволяет эффективно моделировать и изучать однонаправленные нейронные сети и сети с обучением по методу векторного квантования. Нейрочип имеет разрядность по 16 бит для переменных, весов и скалярного произведения (для моделирования передаточной функции) соответственно. Кроме того, 47 бит предназначены для задания пороговых значений и 38 бит используются для выходных данных. В одном корпусе содержится 4 однотипных микропроцессора, которые функционируют параллельно с тактовой частотой 50 МГц. Для реализации систолической архитектуры между отдельными процессорными элементами введена задержка 4 нс. Подобную систолическую цепочку можно расширить путем внешнего соединения нескольких нейрочипов. Для функционирования нейрочипа необходима внешняя память. В [34,63] описана вычислительная плата на основе нейрочипа MA16, выполненная на плате VME9U стандарта VME64. Помимо нейрочипа основными элементами процессорной платы являются девять модулей ППЗУ емкостью 64К каждый, работающих с частотой 83 МГц. С помощью такой процессорной платы можно реализовать двухслойную однонаправленную сеть на 16 входов. Две такие сети расположены на одной ускорительной плате. Время задержки в однослойной сети составляет 3 мкс. Такая сеть реализует дискриминатор Фишера. С помощью трехслойной сети, состоящей из пяти скрытых нейронов и одного

выходного нейрона, решалась аналогичная задача. При этом получены идентичные результаты. Авторами получены также предварительные результаты, позволяющие оценить возможности аналогового нейронного чипа ETANN и цифрового нейропроцессора MA16. Как и следовало ожидать, положительным качеством цифровых нейронных чипов является высокая стабильность их работы в зависимости от температуры.

Цифровой нейронный чип TOTEM. Особенностью данного нейронного чипа является то, что он предназначен для аппаратной реализации так называемого алгоритма RTS (Reaction Tabu Search — реакция на нахождение запрета). Вначале рассмотрим суть RTS-алгоритма. Одним из существенных недостатков многослойных ИНС является то, что процесс их обучения, как уже отмечалось, является весьма трудоемким. Другие проблемы проявляются в том, что нередко трудно интерпретировать внутренние представления многослойного перцептрона. Так, для данного входного вектора выходное значение получается путем нелинейной комбинации входных значений с вовлечением в процесс внутреннего слоя, который чаще всего трудно контролировать. Отчасти это связано с тем, что алгоритмы настройки ИНС, подобные методу обратного функционирования и требующие выполнения операции дифференцирования, имеют ряд недостатков, ограничивающих их применение. Так, после нахождения первого же локального минимума работа алгоритма прекращается. Практически это означает, что можно начинать обучение от новой произвольной точки, и предыдущие результаты оказываются бесполезными. Кроме того, операция дифференцирования требует значительных аппаратных затрат и времени. Предъявляются высокие требования к точности представления весов. И, наконец, в процессе обучения необходимы все данные, получаемые на выходном слое. Причем поиск ведется в локальной области. Существуют несколько направлений в создании моделей ИНС, позволяющих существенно упростить и ускорить процесс обучения нейронной сети. Ниже будут рассмотрены два таких направления, которые реализованы в виде нейронных чипов.

Вначале рассмотрим алгоритм работы нейронного чипа. Дабы избежать названных выше трудностей и упростить решение проблемы обучения, в [64] предложен довольно простой алгоритм обучения ИНС, который радикально отличается от широко известных и относительно просто реализуется с помощью специализированного нейронного чипа. Основное отличие заключается в том, что задача обучения сети сводится к задаче оптимизации методами комбинаторного анализа. По своей сути RTS-алгоритм является алгоритмом локального поиска. Он оптимизирует заданную оценочную функцию E путем использования итеративных «сеточных» компонент для смещения поиска в направлении меньших значений E . С целью исключения бесконечных циклов используется стратегия запретов (tabu). При этом веса n , содержащие b бит, представляются в виде цепочки длиной $l = nb$. Причем значения весов отличаются относительно друг друга всего лишь на один бит. Другими словами, двоич-

ные веса представляют собой слова кода Грэя. В свою очередь, эта цепочка рассматривается в качестве угла куба в l -м пространстве, и задача заключается в поиске угла с минимальной ошибкой. Один бит задает направление перехода (вращения) от одного угла к другому. Далее, искомая траектория $X^{(t)}$ генерируется в выбранной плоскости Z , и последовательность точек X выбирается непосредственно среди соседних точек, которые получаются как следствие элементарных вращений $m_i (i = 1, \dots, l)$. Причем на каждом шаге итеративного процесса в результате элементарных вращений выбираются наименьшие значения оценочной функции E . По мере выполнения элементарных вращений, вращения в обратном направлении запрещаются путем введения параметра T . Другими словами, с помощью параметра T определяется общее количество итераций. Первоначальное значение $T = 1$, и его текущее значение может увеличиваться или уменьшаться в зависимости от успеха в процессе обучения. Применение RTS-алгоритма для обучения ИНС подробно описано в [65].

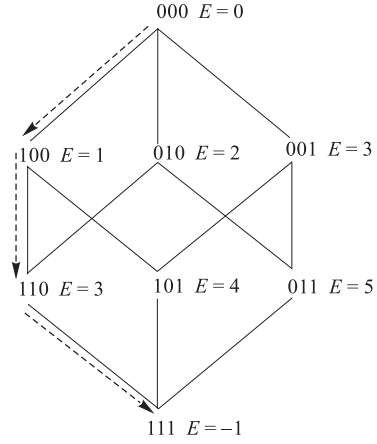


Рис. 25. Пример, поясняющий работу RTS-алгоритма. Обучаемая система достигает глобального минимума $f^{(3)} = [1, 1, 1]$ с оценочной функцией $E^{(3)} = -1$. Стрелки указывают направления вращения

Рассмотрим простой пример (рис. 25). Пусть изучаемая поверхность F представляет собой трехбитовую последовательность ($f = [b_1, b_2, b_3]$), и оценочная функция имеет вид

$$E([b_1, b_2, b_3]) = b_1 + 2b_2 + 3b_3 - 7b_1b_2b_3.$$

На рис. 25 приведены возможные точки (вершины трехмерного двоичного куба) вместе с оценочной функцией. Точка $f^{(0)} = [0, 0, 0]$ с $E(f^{(0)}) = 0$ представляет собой локальный минимум, поскольку всевозможные вращения дают большее значение оценочной функции. Наиболее приемлемым вращением является вращение m_1 , так что в результате имеем $f^{(1)} = [1, 0, 0]$. Следует отметить, что вращение выполняется, даже если $E(f^{(1)}) = 1 \geq E(f^{(0)})$. Поэтому система «покидает» локальный минимум. Если $T^{(1)} = 0$, то более благоприятным вращением из точки $f^{(1)}$ будет опять-таки m_1 , и система возвратится в стартовую точку. Если же значение $T^{(t)}$ остается равным нулю, то обучаемый объект будет повторять ограниченный цикл $[0, 0, 0] \rightarrow [1, 0, 0] \rightarrow [1, 0, 0] \dots$. С другой стороны, если $T^{(t)} = 1$, то при $t = 1$ вращение m_1 запрещено, поскольку оно уже использовалось. Поэтому приемлемым является

ход в направлении m_2 . В результате получаем $f^{(2)} = [1, 1, 0]$, и оценочная функция равна $E(f^{(2)}) = 3$. И, наконец, при $t = 2$ приемлемым вращением является m_3 , хотя бы потому, что оно еще не использовалось. При этом система достигает глобального минимума: $f^{(3)} = [1, 1, 1]$ с оценочной функцией $E^{(3)} = -1$.

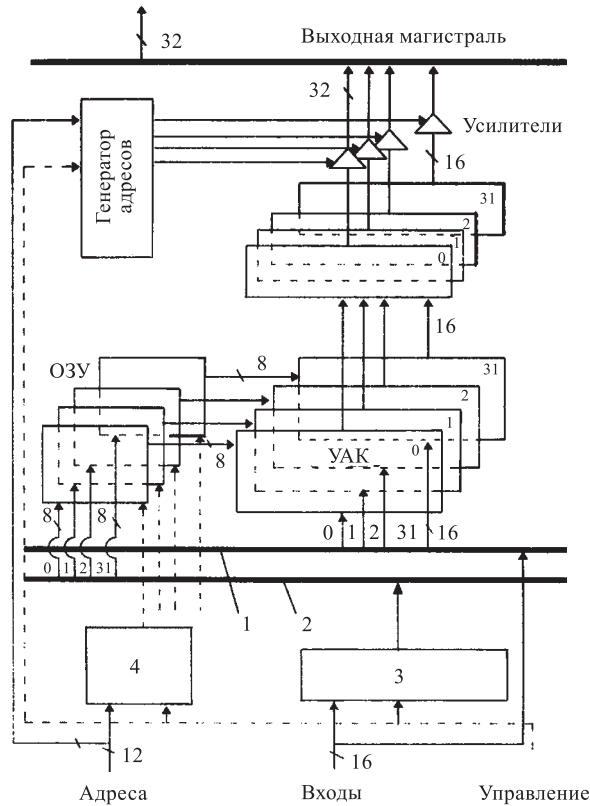


Рис. 26. Блок-схема нейрочипа TOTEM: УАК — умножитель-аккумулятор; 1, 2 — магистрали; 3 — преобразователь кода Грэя в двоичный код; 4 — генератор адресов весовых коэффициентов. Нейрочип содержит 32 процессорных элемента. Каждый процессор вычисляет скалярное произведение весовых коэффициентов w_{ij} и последовательности x_i , которые в режиме «от одного на все» поступают от входной магистрали. С целью ускорения выполнения операций используется конвейерный способ выполнения операций умножения со сложением. Это значит, что после трехцикловой задержки на выходах нейрочипа с циклом 35 нс выполняется до 32 умножений со сложением. Сигмоидная функция реализована методами табличной арифметики

На рис. 26 приведена блок-схема модуля [66]. С целью оптимизации реализации RTS-алгоритма нейрочип TOTEM выполняет только операции, характерные для данного алгоритма, и поэтому он представляет собой копроцессор к процессору общего назначения. Модуль содержит 32 процессорных элемента, работающих параллельно. В свою очередь, каждый процессор P_i в основном вычисляет скалярное произведение весов w_{ij} , которые хранятся в ОЗУ, и последовательности x_i , которые (последовательно, но в режиме «от одного на все» (broadcast)) поступают от входной магистрали. Для временного хранения скалярных произведений имеются регистры, данные от которых могут быть снова переданы на входную магистраль. В процессоре используется динамическая память с последовательной адресацией. С целью ускорения выполнения операций в модуле используется конвейерный способ выполнения операций умножения со сложением. Это означает, что после первоначальной трехцикловой задержки на выходах процессора с циклом 35 нс получается до 32 умножений со сложением. Сигмоидная функция реализована с помощью табличной арифметики. Нейрочип TOTEM выполнен по стандартной КМОП-технологии с шириной дорожки 1,2 мкм и содержит 250 тыс. транзисторов. В следующей версии будет использована субмикронная технология (0,8 мкм). Корпус микросхемы содержит 132 вывода. Поскольку нейрочип представляет собой копроцессор к более универсальному процессору, то для построения ИНС разработана процессорная плата, совместимая с персональными компьютерами. Области применения нейрочипа: классификация объектов, распознавание образов в экспериментах по физике высоких энергий и в системах контроля. Вопросы конкретного применения нейрочипа TOTEM в эксперименте по поиску гипотетических хиггсовских частиц обсуждаются в [67].

Нейрочип CNAPS-1064 (Connected Network of Adaptive Processors — сеть связанных адаптивных процессоров). Нейрочип содержит 64 процессорных элемента, работающих параллельно с частотой цикла 25 МГц [68]. Процессорный элемент содержит 32-разрядный сумматор, умножитель 16×8 , 16-разрядное логическое устройство, 4 Кбайт локальной памяти с устройством адресации, регистр, внутренние и внешние магистрали для обмена 16-разрядными адресами и 8-разрядными данными. Архитектура нейрочипа позволяет реализовать матричный процессор, содержащий до 512 аналогичных нейрочипов. Фирма-разработчик ASI поставляет также процессорную плату, выполненную в стандарте VME и содержащую четыре нейрочипа CNAPS-1064, 16 Мбайт файлового ОЗУ, $64\text{К} \times 64$ бит программной памяти, контроллеры и интерфейс ввода-вывода, выполненный в виде навесной конструкции, которая сопрягается с основной нейроплатой. В [69] описана система обработки аэрокосмических изображений, в которой используется описываемый нейрочип. В [68] нейрочип используется в эксперименте по физике высоких энергий для фильтрации данных, регистрируемых с помощью уста-

новки Н1. Триггерная система содержит четыре уровня фильтрации физических событий $L1 \div L4$ с временем решения 2, 20, 800 и 100000 мкс соответственно. Для фильтрации событий на втором уровне с временем решения 20 мкс создана мощная вычислительная нейросистема на базе стандартной нейроплаты фирмы ASI, которая наряду с другими обычными процессорными модулями составляет основу триггерной системы второго уровня.

Эксперименты с нейрочипом ZISC036 [70]. Так же, как и в предыдущем случае, процесс обучения нейрочипа носит локальный характер. В данном модуле реализован так называемый RBF-алгоритм (Radial Base Function). Кратко суть алгоритма заключается в следующем. RBF-алгоритм [8,18] базируется на простой интуитивной идее, которая заключается в том, что произвольная функция $y(x)$ может быть аппроксимирована с помощью линейной суперпозиции системы локализованных базовых функций $h_j(x)$. Такой подход, по существу, приводит к модели ИНС, близкой по функциональным возможностям к многослойным ИНС. Новый тип ИНС описывается с помощью выражения [8]:

$$y_k(\mathbf{x}) = \sum_{j=0}^m w_{kj} h_j(\mathbf{x}),$$

где m — количество базовых функций; значение w_{kj} определяется из системы линейных уравнений. Функция $h_j(\mathbf{x})$ представляет собой функцию активации j -го нейрона скрытого слоя и \mathbf{x} — входной вектор. В отличие от однонаправленной сети в данной сети вводится также смещение в выходные нейроны (см. блок-схему нейронной сети, реализующую данную модель [8]).

Обучение сети выполняется в два этапа. Вначале оптимизируются параметры базовой функции и затем вычисляются веса. Высокое быстродействие ИНС, в которой реализуются радиально-базовые функции, обуславливается тем, что на выходе каждого нейрона получается значение функции, отличающееся от нуля в ограниченной области пространства. Другими словами, процесс обучения носит локальный характер, и поэтому не требуется комплексная информация от других нейронов, что существенно упрощает их структуру. Подобные сети эффективно решают классификационные проблемы. Возможно несколько модификаций алгоритмов, использующих радиально-базовые функции. Одна из них реализована в нейрочипе ZISC036, который разработан фирмой IBM и описан в [70]. Он состоит из 36 логических ячеек и блока принятия решений. Каждый нейрон имеет по два входа: один вход используется для записи контрольного вектора, а на второй вход подаются данные, характеризующие область влияния данного нейрона. Принцип работы нейрочипа можно представить, рассмотрев процесс обучения ИНС.

- Каждый нейрон параллельно вычисляет так называемое блоковое кодовое расстояние Манхеттена d между входным вектором \mathbf{V} и вектором прототипа \mathbf{P} :

$$d = \sum (v_i - p_i),$$

или выбирает максимальное значение

$$|v_i - p_i|, \quad i = 1, \dots, n$$

для n -элементного вектора.

- Каждый прототип имеет область влияния q такую, что если $d < q$, то нейрон включается.
- В случае, когда все прототипы удовлетворяют определенным условиям, вырабатывается признак правильной идентификации между входными данными и прототипами.
- Если срабатывают несколько нейронов, вырабатывается признак неопределенности в классификации.
- И, наконец, в случае отсутствия положительного результата, т.е. когда ни один из нейронов не сработал, имеет место неопределенность в состоянии нейросети.

Нейрочип ZISC036 выполняет до 250 тыс. операций в секунду над векторами и прототипами, что эквивалентно $2 \cdot 10^9$ обычных операций в секунду. На плате VME монтируется до четырех нейрочипов с возможностью каскадирования до сорока модулей. Данный нейрочип предполагается использовать в эксперименте по поиску тяжелых частиц Хиггса [71]. С этой целью было проведено моделирование, результат которого приведен на рис. 27. Здесь показано двумерное распределение импульсов частиц, соответствующих полезным сигналам и шумам, которые, как видно из рисунка, разделяются. В качестве входных сигналов использовались импульсы p и поперечные импульсы p_t четырех различных частиц (общее число входов сети равнялось восьми). Сигналы исходили или из модели реакции распада тяжелых хиггс-частиц или от шумовых процессов. Полученные результаты оказались более высокими по сравнению с использованием для аналогичных целей нейрочипов ETANN и CLN32/64.

Известен еще один нейрочип — SANDY/1, в котором используется модель радиально-базовых функций [72]. Нейрочип SANDY/1 представляет собой каскадируемую систолическую матрицу, которая может быстро моделировать многослойную прямооточную сеть, реализующую радиально-базовые функции и некоторые другие модели. Такая архитектура существенно расширяет области применения нейрочипа. Нейрочип SANDY/1 вычисляет выражение (1) с помощью n процессорных элементов, если вектор x имеет раз-

мерность n . Это возможно, поскольку все компоненты x_i вычисляются независимо. Как отмечается в [72], одним из недостатков матрично-векторного умножения является то, что количество циклов, необходимых для передачи элементов матрицы W , больше, чем активационных векторов v_i . В результате получаются пустые циклы.

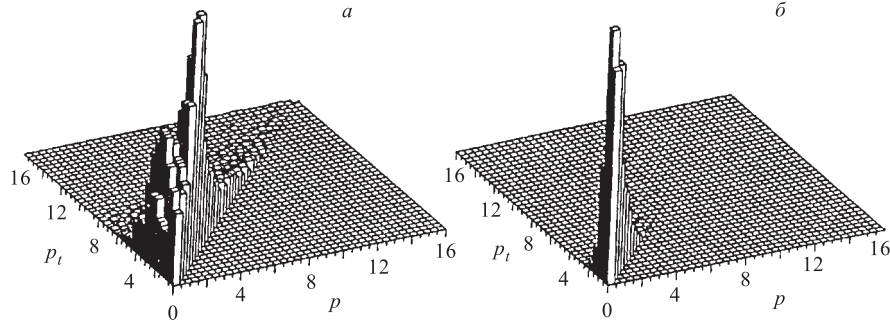


Рис. 27. Образцы распределений сигналов, используемых для тренировки ИНС: а) полезные сигналы; б) шумы; p — импульсы частиц; p_t — поперечные импульсы частиц. Видно, что картинки, иллюстрирующие распознавание полезных сигналов и шумов, существенно отличаются [71]

Количество циклов можно существенно уменьшить, если выполнять процедуру умножения двух матриц, что и реализовано в архитектуре нейрочипа. В этом случае четыре активационных вектора в процессе распознавания вычисляются за четыре цикла ($n = 4$), и результат операции передается в следующий процессорный элемент. 40-разрядная сетка АЛУ позволяет обчислить до 512 нейронов без потери необходимой точности. С целью обеспечения совместимости с 16-разрядной магистралью и упрощения каскадирования отдельных нейрочипов введен блок дискриминации 40-битных значений в 16-битные. Для генерации функции активности имеется отдельная микросхема. Для реализации не только алгоритма радиально-базовых функций, но и алгоритма Кохонена в нейрочипе имеются два дополнительных блока: дополнительное АЛУ и блок определения наиболее активного нейрона.

В нейрочипе SANDY/1 реализуется представление радиально-базовых функций, требующее вычисления евклидова расстояния между векторами w и v , представляющих собой весовой вектор и вектор активности соответственно. Эта процедура требует вычисления корня квадратного из суммы разностей квадратов. Для этой цели в нейрочипе имеются два дополнительных устройства: 16-разрядный сумматор для вычисления разностей $w_i - v_i$, работающий в конвейерном режиме с основным АЛУ, и АЛУ для вычисления сумм квадратов. Квадратный корень находится табличным методом. Нейро-

чип SANDY/1 выполнен по КМОП-технологии с шириной дорожки 0,8 мкм в корпусе, содержащем 144 вывода. При цикле 20 нс выполняется $2 \cdot 10^8$ операций в секунду. В табл. 2 приведены сравнительные характеристики трех вышеописанных цифровых модулей.

Таблица 2. Сравнительные характеристики нейрочипов

Нейрочип	Входной вектор: количество бит, частота поступления	Время вычисления	Задержка	Структура ИНС
CNAPS	8 бит, 20 МГц	8 мкс — одно событие	—	$64 \times 64 \times 1$
SANDY/1	16 бит, 40 МГц	20,5 мкс — четыре события	27	$64 \times 64 \times 1$
MA16	16 бит, 40 МГц	5,5 мкс	8	$16 \times 5 \times 1$
SANDY/1	16 бит, 40 МГц	2 мкс — четыре события	3,6	$16 \times 5 \times 1$

Как видно из табл. 2, нейрочип SANDY/1 имеет более высокую производительность, нежели нейрочипы CNAPS и MA16. Так, с помощью нейрочипа SANDY/1 одно событие обрабатывается примерно за 5 мкс, с помощью нейрочипа CNAPS на обработку одного события требуется 8 мкс при одинаковой структуре ИНС $64 \times 64 \times 1$. При одинаковой структуре ИНС $16 \times 5 \times 1$ для обработки одного события с помощью нейрочипа SANDY/1 требуется времени в пять раз меньше, нежели с помощью нейрочипа MA16. Данная разработка предназначена для использования в системах быстрого отбора событий в экспериментах по физике высоких энергий.

Применение активной перепрограммируемой памяти. Концепция активной перепрограммируемой (АПП) памяти базируется на использовании размещенных в одном корпусе микросхемы перепрограммируемой логической матрицы (ПЛИМ) и доступных загрузчику файловых регистров большой емкости (банки памяти). Это направление в цифровой технике активно развивается фирмой «Xilinx» с начала 90-х годов [73]. Чаще всего на базе такой технологии создаются быстрые копроцессоры, выполняющие специфические функции. Один из таких копроцессоров, предназначенный для быстрой обработки данных, которые будут регистрироваться в экспериментах на большом адронном коллайдере, описан в [74]. Процесс разработки и контроля разработки может выполняться квалифицированным инженером на рабочем месте при наличии соответствующих программных и аппаратных средств. При этом цикл создания сложного устройства обычно не превышает одной недели и менее. Будучи однажды запрограммированным, устройство затем работает как обычная большая микросхема до выключения источника питания. Логика ра-

боты копроцессора описывается с помощью языка высокого уровня. Данные в модуль загружаются в последовательном коде или байтами. Ожидается, что к 2001 г. в одном корпусе может содержаться до 25К неоднократно перепрограммируемых вентилях с тактовой частотой 200 МГц. Широкое практическое применение концепция АПП нашла благодаря разработке процессорной платы DecPeRLe-1 [61] фирмой DEC, структурная схема которой кратко описана в [74,75]. На одной плате расположено 23 модуля фирмы «Xilinx» и четыре банка памяти емкостью 1 Мбайт каждый, которые объединены с помощью внутренних магистралей. Основу копроцессора составляет вычислительная матрица M , состоящая из 16 модулей ПЛМ. В свою очередь, каждый модуль имеет связи с соседними. В дополнение к этому копроцессор с помощью мультиплексоров, которые представляют собой отдельные модули, имеет доступ к каждой из четырех 64-разрядных магистралей и к банкам памяти. С помощью других модулей осуществляется связь с основным процессором и генерация адресов для АПП. Как показано в [75], копроцессор эффективно используется в прототипе системы сбора и обработки физических данных установки ATLAS в качестве классификатора частиц. Была использована модель однонаправленной сети по схеме 12-6-4. Три выхода используются для классификации частиц, например, «электрон», «мюон» или «струя» из частиц. Четвертый выход является резервным. Суть проблемы заключается в том, что в будущем эксперименте на установке ATLAS, субдетекторы которой будут содержать несколько миллионов каналов регистрации, требуется быстро (менее чем за 2 мкс) определить те области детекторов, в которых имеется признак полезного события (область интересов), обработать эти данные и принять решение о передаче отфильтрованной информации на второй уровень триггерной системы. Эта задача решается с помощью логических процессоров. На втором уровне триггерной системы, где время решения должно составлять порядка 10 мкс, выполняется классификация зарегистрированных частиц, по которой можно судить о наличии полезного события. И, наконец, на третьем уровне воссоздается топология полезного события. Использование ИНС в такой установке позволит существенно упростить триггерную систему и повысить качество регистрации полезных событий.

Разработка быстрого специализированного нейрокомпьютера. Как уже отмечалось выше, ведутся интенсивные исследования, направленные на использование в системах запуска физических установок быстродействующих ИНС и нейрокомпьютеров. В [76] описан специализированный нейрокомпьютер, с помощью которого определяется множественность и координаты треков зарегистрированных частиц в эксперименте CPLEAR за время 75 нс. Целью эксперимента является прецизионное измерение нарушения CP-симметрии путем использования первичных K^0 - и \bar{K}^0 -частиц от распада антипротонов в водородной мишени.

Прежде чем описывать архитектуру ИНС, кратко рассмотрим параметры экспериментальной установки. Ядром установки является цилиндрический детектор, состоящий из нескольких проволочных камер, расположенных вокруг водородной мишени. Ближе к мишени расположены два слоя пропорциональных камер, шесть слоев дрейфовых камер, два слоя стримерных трубок, сцинтилляционный годоскоп, черенковские счетчики и покрывающий все детекторы газонаполненный калориметр. Информация от трековых детекторов используется в нейροкомпьютере для подсчета количества треков (два или четыре) и их положения. Исходя из практических соображений, каналы регистрации трековых детекторов разделены на 64 сектора. Одна процессорная плата, представляющая собой локальную сеть, используется для обработки данных, поступающих от пяти соседних секторов. В одном VME-крейте располагается до 16 процессорных плат. Для построения триггерной системы потребовалось четыре аналогичных крейта. Все процессорные платы работают по одному и тому же алгоритму. На рис. 28 приведена блок-схема

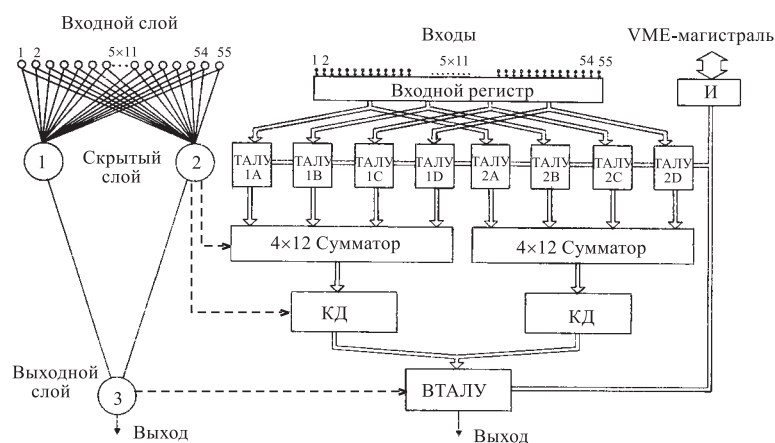


Рис. 28. Блок-схема одной платы специализированного нейροкомпьютера. ТАЛУ1А÷ТАЛУ1D, ТАЛУ2А÷ТАЛУ2D — АЛУ с табличным методом решения; ВТАЛУ — выходное ТАЛУ; КД — компрессор данных; И — элементы И. На входной слой ИНС поступают данные от детектора. На входном регистре запоминаются желаемые образцы. Пунктирными линиями условно показаны связи между нейронной сетью и остальной частью нейροкомпьютера

одной процессорной платы. ИНС представляет собой однонаправленную 55-2-1 сеть, которая, с целью получения высокого быстродействия, выполнена на обычных микросхемах и с широким использованием табличных методов решения, позволяющих быстро вычислять сигмоидную функцию и

производить умножение входных значений входных сигналов на весовые коэффициенты. Для этих целей используются блоки ТАЛУ1А÷ТАЛУ1D и ТАЛУ2А÷ТАЛУ2D. Распараллеливание блоков ТАЛУ обеспечивает меньшую емкость табличной памяти и в конечном итоге более высокое быстродействие. В качестве адресов этих блоков служат значения сумм входного образца, умноженные на соответствующие весовые коэффициенты. Выходы блоков ТАЛУ подключены к входам быстрых сумматоров. На рисунке не показан еще один блок ТАЛУ, который подключен к нейронам 1 и 2 скрытого слоя. В случае выработки сигнала о регистрации полезного события данные передаются для дальнейшей обработки. Отмечается, что разработанный нейрокompьютер хорошо адаптируется к изменениям измеряемых параметров. В цитируемой работе проводится сравнение неэффективности отбора двух и четырех треков, выполненное с помощью обычного аппаратного устройства, с помощью программных средств и с помощью ИНС. Получены следующие величины неэффективности отбора: 0,218 и 0,444; 0,083 и 0,217; 0,144 и 0,296.

11. ЗАКЛЮЧЕНИЕ

Нейрочипы и создаваемые на их основе искусственные нейронные сети могут быть успешно использованы для быстрого отбора событий в экспериментах по физике высоких и сверхвысоких энергий, для классификации элементарных частиц и в других направлениях научного приборостроения. Этому способствует быстрое развитие математических алгоритмов ИНС и достижений полупроводниковой технологии. Широкое развитие получили вычислительные нейроплаты и создаваемые на их основе нейрокрейты в стандарте VME. В равной степени используются аналоговые, аналогово-цифровые и чисто цифровые ИНС и нейрокompьютеры. Наиболее широкое применение находят трехслойные однонаправленные ИНС.

Характерно также, что нейрочипы и создаваемые на их основе ИНС изготавливаются не только промышленным способом, но и силами разработчиков научной аппаратуры путем использования такого перспективного направления, как неоднократно электрически перепрограммируемые большие интегральные микросхемы [77]. Используя язык высокого уровня и персональную ЭВМ, квалифицированный разработчик может в течение нескольких дней изготовить необходимый нейрочип. Как показано в [78], наиболее перспективными для использования в научных исследованиях и в приборостроении являются четыре нейрочипа: ETANN, ZISC036, TOTEM и CNAPS.

Приложение. Характеристики рассматриваемых нейрочипов

Нейро-чип	Фирма	Аналоговый цифровой	Общие характеристики
ETANN	Intel	Аналоговый	В корпусе микросхемы, содержащей 208 выводов, имеется 64 нейрона, содержащих 10240 синапсов. Максимально возможное число входных сигналов — 128 на один нейрон. 2×64 выводов предназначены для входных сигналов, и 64 вывода являются выходными, которые можно соединять с входными выводами. Задержка сигналов в одном модуле составляет 1–3 мкс [29].
ANNA	AT&T Bell Laboratories	Аналогово-цифровой	Количество синапсов 4096. Количество элементов смещения 256. Одновременно выполняет до 2 млн. операций умножений и сложений в секунду. Типичная область применения — распознавание рукописных текстов. Данные в цифровой форме поступают на входы с тактовой частотой 20 МГц в последовательном коде [30,31].
CLN32	Bellcore	Аналогово-цифровой	Кроме 32 нейронов и 496 адаптивных синапсов на кристалле имеется встроенное устройство для обучения. Весовые коэффициенты запоминаются в цифровой форме. С помощью дополнительного чипа CLN64 можно подключить еще 1024 адаптивных синапса. Задержка сигналов в однослойной сети равна 1 мкс [45].
MESA	MESA	Аналогово-цифровой	Архитектура нейрочипа позволяет создавать однонаправленную сеть, содержащую 70 входных нейронов, 6 скрытых нейронов и несколько выходов. Это достигается тем, что взвешенные выходы каждого нейрона скрытого слоя имеют внешние выводы. В то время как входные сигналы являются аналоговыми, пятиразрядные синаптические веса программируются от внешнего процессора [46].
L-Neuro	Telmat	Цифровой	Базовый нейрочип, предназначенный для создания современных нейрокомпьютеров под управлением транспьютеров. Особенности нейрочипа являются аппаратная реализация параллельного выполнения базовых операций в процессе обучения и распознавания. Локальная память весовых коэффициентов составляет 1 кбайт [55].
НТЦ «Модуль»	НТЦ «Модуль»	Цифровой	Нейрочип отечественного производства предназначен для построения сложных нейросетей. В основу архитектуры нейрочипа положен оригинальный метод выполнения его базовых операций с помощью однородной вычислительной среды, которая позволяет производить операции произвольной разрядности (от 1 до 64 бит). При тактовой частоте 30 МГц нейрочип выполняет до 720 миллионов операций с плавающей точкой в секунду [60].
SYNAPS1 N110	Siemens	Цифровой	Нейрокомпьютер, реализованный на базе восьми нейрочипов MA16. Нейрокомпьютер выполняет до 3,2 млрд. операций умножения (16×16 и сложений в секунду). Емкость памяти 32 Мбайт [61].
MA16	Siemens	Цифровой	Работает с тактовой частотой 50 МГц в систолическом режиме. Необходима внешняя память. Точность представления данных 16 бит [63].
TOTEM	INFN Universita di Trento.	Цифровой	Реализует RTS-алгоритм и содержит 32 процессорных элемента (ПЭ), работающих в конвейерном режиме. Используется в качестве копроцессора к персональной ЭВМ. Выполняет до 32 умножений со сложением за 35 нс. Емкость памяти весовых коэффициентов 32 кбит [66,67].
SNAPS-1064	Adaptive Solutions	Цифровой	Содержит 64 ПЭ, работающих параллельно с тактовой частотой 20 МГц. Архитектура нейрочипа позволяет создавать матричный процессор, содержащий 512 ПЭ. Память весовых коэффициентов равна 4 кбайт.
ZISC036	IBM	Цифровой	Реализует алгоритм радиально-базовых функций (РБФ). Нейрочип содержит 36 идентичных нейронов, подключенных к общей магистрали и управляющее устройство. Используется в качестве копроцессора для быстрого вычисления весовых коэффициентов. Тактовая частота 20 МГц [70,71].
SANDY/1	Research center Karlsruhe	Цифровой	Реализует алгоритм РБФ и многослойный перцептрон. Имеется четыре процессорных элемента, работающих в систолическом режиме с возможностью каскадирования. Память весовых коэффициентов равна 64×16 . Выполняет 600 млн. операций в секунду [72].

СПИСОК ЛИТЕРАТУРЫ

1. **Фишбах Д.Д.** — В мире науки, 1992, № 11,12, с.10.
2. **Хингтон Д.Е.** — В мире науки, 1992, № 11,12, с.103.
3. **Кэмп Д.В.** — В мире науки, 1992, № 11,12, с.200.
4. **Kohonen T.** — Neural Networks, 1988, v.1, No.1, p.3.
5. **Кохонен Т.** — Ассоциативные запоминающие устройства. Ред. Зуева В.И. М.: Мир, 1982, 375 с.
6. **Ачасова С.М.** — Программирование, 1991, № 2, с.40.
7. **Lippman R.P.** — IEEE ASSP Magazine. April 1987, p.4.
8. **Bishop C.M.** — Review Science Instruments, 1994, v.65, No.6, p.1803.
9. **Галушкин А.И.** — Зарубежная радиоэлектроника, 1995, № 5, с.4.
10. **Галушкин А.И.** — Зарубежная радиоэлектроника, 1995, № 6, с.4.
11. **Кисель И.В., Нескоромный В.Н., Ососков Г.А.** — Физика элементарных частиц и атомного ядра, 1993, т.24, вып.6, с.1551.
12. **Галушкин А.И., Иванов В.В., Картамышев М.Г.** — Зарубежная радиоэлектроника, 1997, № 2, с.3.
13. **Логовский А.С., Якушев Д.Ж.** — Зарубежная радиоэлектроника, 1997, № 2, с.11.
14. **Кирсанов Д.В.** — Зарубежная радиоэлектроника, 1997, № 2, с.19.
15. **Степанов М.В.** — Зарубежная радиоэлектроника, 1997, № 2, с.32.
16. **Bortolotto C., De Angelis A., De Groot N. et al.** — Preprint INFN/AE-92/13. Frascati, 1992.
17. **Peterson C.E.** — Nucl. Instrum. and Methods, 1989, v.A279, No.3, p.537.
18. **Renalds S.** — Electronics Letters, 1989, v.25, No.7, p.437.
19. **Amendolia S.R.** — 1992 CERN School of Computing. CERN 93-03. Geneva, 1993, p.135.
20. **Fisher R.A.** — Annals Engenics, 1936, v.7, p.179.
21. **Mazzani P., Odario R.** — Z. fur Physik (C), 1993, v.59, No.2, p.273.
22. **Baldanza C., Bisi F., Cotto-Romansino A. et al.** — Nucl. Instrum. and Methods, 1995, v.A361, No.3, p.506.
23. **Розенблаг Ф.** — Принципы нейродинамики. М.: Мир, 1965, 480 с.
24. **Фомин Д.В.** — Приборостроение, 1995, № 1,2, с.57.
25. **Bursky D.** — Electronic Design, 1997, v.45, No.25, p.67.
26. **Кусуль М.Д.** — Управляющие системы и машины, 1993, № 5, с.6.
27. **Ясер С., Псалтис Д.** — В мире науки, 1987, № 5, с.42.
28. **Мануэль Т.** — Электроника, 1988, т.61, № 16, с.58.
29. **Brauch J., Tam S.M., Holler M.A. et al.** — IEEE Micro, 1992, No.6, December, p.34.
30. **Boser V., Sackinger E., Bromley J. et al.** — IEEE Micro, 1992, February, p.32.
31. **Sakinger E., Boser V.E., Bromley J. et al.** — IEEE Transaction on Neural Networks, 1992, v.3, No.3, p.498.
32. **Molnar J., Szekely G., Lindblad Th. et al.** — ICFA Instruments Bulletin, 1993, No.10, p.24.

33. **Lindsey C.S., Lindblad Th., Vollaro J.R. et al.** — Nucl. Instrum. and Methods, 1994, v.A351, No.1,2, p.466.
34. **Baldanza C., Beichter J., Bisi F. et al.** — Nucl. Instrum. and Methods, 1996, v.A373, No.2, p.261.
35. **Ponting P.J.** — In: Proc. Intern. Conf. on New Trends in Data and Signal Processing in Research. ESONE RTD'95. Warsaw, Poland. Sept. 27–29. 1995, p.35.
36. **Lindblad Th., Szekely G.** — Nucl. Instrum. and Methods, 1993, v.A328, No.3, p.603.
37. **Kolonski H.** — Nucl. Instrum. and Methods, 1995, v.A367, No.1-3, p.16.
38. **Denby B., Lindblad Th., Lindsey C.S. et al.** — Nucl. Instrum. and Methods, 1993, v.A335, No.2,3, p.296.
39. **Hornblad P., Lindblad Th., Lindsey C.S. et al.** — Nucl. Instrum. and Methods, 1993, v.A336, No.1,2, p.285.
40. **Wu D.Y., Cambell M.K.** — Nucl. Instrum. and Methods, 1992, v.A317, No.1,2, p.323.
41. **Denby B. Lindsey C.S., Dikson M. et al.** — Nucl. Instrum. and Methods, 1995, v.A356, No.2,3, p.485.
42. **Conway J.S., Looms C.** — Int. J. of Modern Physics C, 1995, v.6, No.4, p.549.
43. **Akkila T., Franke T., Lindblad Th.** — Nucl. Instrum. and Methods, 1993, v.A327, No.2,3, p.566.
44. **Danielsen M., Go A., Jon-And K. et al.** — Nucl. Instrum. and Methods, 1994, v.A350, No.1,2, p.322.
45. **Lindblad Th., Lindsey C.S., Block F. et al.** — Nucl. Instrum. and Methods, 1995, v.A356, No.2,3, p.498.
46. **Tong Tran T.** — Int. J. of Modern Physics C, 1995, v.6, No.4, p.573.
47. **Eide A.** — Nucl. Instrum. and Methods, 1992, v.A317, No.3, p.607.
48. **Lindsey C., Denby B., Haggerty H. et al.** — Nucl. Instrum. and Methods, 1992, v.A317, No.1,2, p.346.
49. **Akkila T., Lindblad Th., Lund-Jensen B.** — Nucl. Instrum. and Methods, 1993, v.A327, No.3, p.573.
50. **Bishop C.M., Haynes P.S., Smith M.E. et al.** — Neural Computation, 1995, v.7, No.1, p.206.
51. **Sandler D.G., Barrett T.K., Painer D.A. et al.** — Nature, 1991, v.351, No.6324, p.300.
52. **Бондарь Е.А., Логовский А.С.** — Зарубежная радиоэлектроника, 1997, № 2, с.57.
53. **Haynes W.J.** — 1989 CERN School of Computing. CERN 90-06. Geneva, 1990, p.179.
54. **Крысанов А.И.** — Зарубежная радиоэлектроника, 1997, № 2, с.27.
55. **Крысанов А.И.** — Зарубежная радиоэлектроника, 1998, № 1, с.18.
56. **Bianchin S., Cosmo G., Garlatt C. et al.** — Preprint INFN/AE-92/14. Frascati, 1992.
57. **Clarke P.L., Granfield R., Crone G.I. et al.** — ATLAS Int. Note 94-01. Geneva, 1994.
58. Sonitech International Inc. 1993–1994. Product Catalog.
59. **Погорильный А.И.** — Приборостроение, 1996, т.39, № 7, с.31.
60. **Виксне П.Е., Фомин Д.В., Черников В.М.** — Приборостроение, 1996, т.39, № 7, с.13.
61. **Истратов А.Ю.** — Зарубежная радиоэлектроника, 1997, № 2, с.24.
62. **Суньюань Гун** — ТИИЭР, 1984, т.72, № 7, с.133.

63. **Baldanza C., Beichter J., Bisi F. et al.** — Int. J. of Modern Physics C, 1995, v.6, No.4, p.567.
64. **Battiti R., Tecchioli G.** — IEEE Transaction on Neural Nets, 1995, v.6, No.5, p.1185.
65. **Batti R., Lee P.** — In: Proc. Fourth Intern. Conf. on Microelectronics for Neural Networks and Fuzzy Systems. Turin, Italy. Sept. 26–28. 1994, p.17.
66. **Anzellotti G., Battiti R., Lazzizzera I. et al.** — Int. J. of Modern Physics C, 1995, v.6, No.4, p.555.
67. **Dusini S., Ferrary F., Lazzizzera I. et al.** — In: Int. Conf. on Computing in High Energy Physics. 1997. Berlin. April 7–11, p.563.
68. **Goldner D., Getta H., Kolander M. et al.** — Int. J. of Modern Physics C, 1995, v.6, No.4, p.541.
69. **David K., Andes James C., Witham C. et al.** — Neural Networks, 1995, v.8, No.7/8, p.1349.
70. **Lindsey C.S., Lindblad Th., Sekhniadze G. et al.** — Int. J. of Modern Physics C, 1995, v.6, No.4, p.579.
71. **Linblad Th., Lindsey C.S., Minerskjold M. et al.** — Nucl. Instrum. and Methods, 1995, v.A357, p.192.
72. **Fisher T., Eppler W., Gemmeke H.** — In: Proc. Second Workshop on Electronics for LHC Experiments. Balatonfured, Hungary, Sept. 23–27, 1996. CERN/LHCC/96-39, p.245.
73. **Brandenburg G., Gorshutz U., Muller K.D.** — IEEE Transaction on Nuclear Science, 1988, v.35, p.213.
74. **Belostoudtsev D., Bertin P., Bock R.C. et al.** — Nucl. Instrum. and Methods, 1995, v.A356, No.2,3, p.457.
75. **Lundheim L., Legrand I., Moll L.** — Int. J. of Modern Physics C, 1996, v.6, No.4, p.561.
76. **Leingruber F.R., Pavlopoulos P., Steincher M. et al.** — Nucl. Instrum. and Methods, 1995, v.A365, No.1, p.198.
77. ALTERA FLEX 8000 handbook. May 1994.
78. **Vlachos S., Wadhwa W., Lindblad Th. et al.** — Nucl. Instrum. and Methods, 1997, v.A385, No.2, p.361.