

194-98



ОБЪЕДИНЕННЫЙ
ИНСТИТУТ
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ

Дубна

P11-98-194

Г.А.Ососков, В.В.Пальчик, Ю.К.Потребенников,
Г.Т.Татишвили, В.Б.Шепелев¹

ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ
ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ
ГЕОМЕТРИЧЕСКОЙ РЕКОНСТРУКЦИИ СОБЫТИЙ,
ЗАРЕГИСТРИРОВАННЫХ В ЭКСПЕРИМЕНТЕ
EXCHARM

Направлено в журнал «Математическое моделирование»

¹Ивановский государственный университет

1998

1 Введение

В работе предлагается комбинированный алгоритм, позволяющий повысить эффективность восстановления событий в эксперименте EXCHARM (описание спектрометра EXCHARM приведено, например, в работе [1]). Речь идет о распознавании 5-трековых событий со вторичной вершиной в распадной области детектора (см.рис.1).

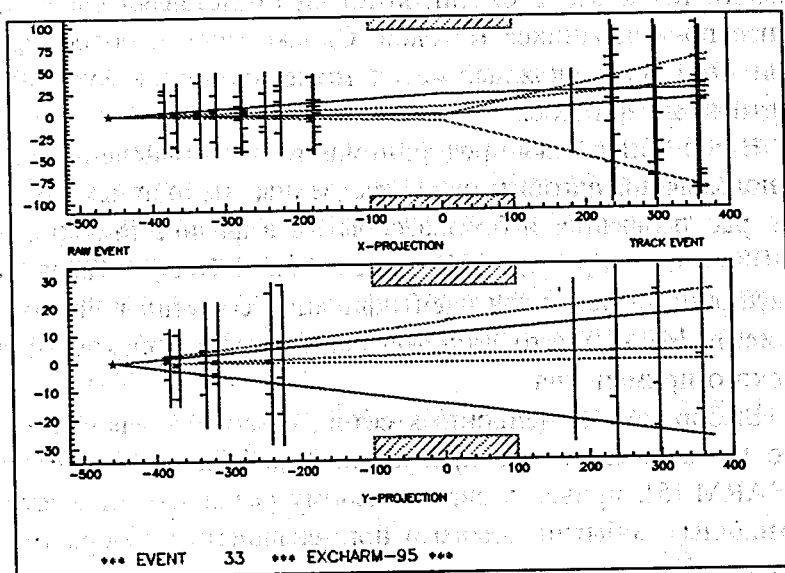


Рис. 1: Типичное событие EXCHARM с фоновыми отсчетами.

В данном эксперименте в качестве детектора используется магнитный спектрометр, расположенный в нейтронном канале 5Н серпуховского ускорителя. В состав детектора входят спектрометрический магнит и многопроволочные пропорциональные камеры, установленные до и после магнита. Точки измерения, ис-

пользуемые непосредственно для распознавания треков, представляются как координаты сработавших проволок в камерах. В связи с тем, что эффективность срабатывания проволок в большинстве камер экспериментальной установки составляет 80-90%, а число фоновых отсчетов сравнимо с числом полезных, стандартные методы восстановления треков [2] становятся неэффективными. В таких условиях появляется необходимость повышения эффективности традиционных методов или поиска более эффективных новых подходов для восстановления треков. При этом должно быть выполнено условие: новые алгоритмы должны распознавать тот класс событий, который представлял сложности для ранее применявшихся методов. Однако новый, более эффективный, но и более сложный метод, может оказаться медленнее стандартных алгоритмов.

В последние годы предпринимаются многочисленные попытки использования алгоритмов искусственных нейронных сетей (ИНС) для распознавания и отбора событий в физике высоких энергий (ФВЭ) (см., например, [3], раздел Neural Nets). Такая попытка предпринималась и для идентификации событий в данном эксперименте [4] на чисто теоретическом уровне, далеко от практического применения.

Выбор метода нейронных сетей, резко отличающегося от ранее применявшегося в программе обработки эксперимента EXCHARM [5], привел к значительному различию множеств нераспознанных событий нового и применявшегося в обработке алгоритмов. Это дало возможность скомбинировать оба метода так, чтобы существенно повысить эффективность распознавания событий в данном эксперименте.

2 Описание нейронной сети

Для распознавания траекторий заряженных частиц был разработан алгоритм, основанный на применении нейронной сети типа Хопфилда [6]. Это полносвязная сеть, являющаяся системой про-

стых вычислительных элементов (нейронов) s_i , связи между которыми характеризуются так называемыми синаптическими весами w_{ij} . Нейроны бинарны, т.е. могут принимать одно из двух возможных состояний. Благодаря связям, смена состояний даже одного из нейронов вызывает эволюцию всей сети, которая в конце концов затухает в некотором равновесном стационарном положении. Хопфилду удалось связать условие выхода сети в стационарное состояние с поведением ее энергетического функционала, который, грубо говоря, может быть представлен как

$$E(s) = -\frac{1}{2} \sum_{ij} s_i w_{ij} s_j. \quad (1)$$

Он доказал теорему, по которой любое стационарное состояние сети с симметричными весами $w_{ij} = w_{ji}$, $w_{ii} = 0$ соответствует одному из локальных минимумов энергетического функционала вида (1). Эта теорема привела авторов работ [7, 8] к идее об отображении данных траекторных измерений на множество состояний нейронной сети, в которой синаптические веса должны быть определены так, чтобы в процессе эволюции сети поощрять связи между нейронами, соответствующими измеренным трекам, а ненужные связи – подавлять.

Для этого энергетический функционал в работах [7, 8] был определен как состоящий из двух частей:

$$E = E_{cost} + E_{constraint}. \quad (2)$$

Так как распознавание ведется для гладких треков, представляемых из-за дискретизации данных в виде наборов сегментов прямых линий, то первый стоимостной член выбирается так, чтобы он поощрял *короткие смежные сегменты с малым углом* между ними. В реальных условиях нужно учитывать неэффективность камер, что приводит к отсутствию некоторых точек на треке, а следовательно, и к удлинению сегментов. Вводя нейроны s_{ij} , отвечающие сегменту, соединяющему экспериментальные точки i и j , получим стоимостной член энергетической функции в виде:

$$E_{cost} = \begin{cases} -\frac{1}{2} \sum_{ijkl} \delta_{jk} \frac{\cos^m \theta_{ijl}}{r_{ij} + r_{jl}} s_{ij} s_{kl}, & \text{если } \theta_{ijl} \leq \alpha_{con}, \\ 0 & \text{, если } \theta_{ijl} > \alpha_{con}, \end{cases} \quad (3)$$

где θ_{ijl} – угол между сегментами, соединяющими точки i с j и j с l , α_{con} – константа, вычисляемая с учетом геометрических особенностей детектора, а m – целочисленный показатель степени (обычно подбирается вручную в зависимости от используемых данных измерений).

Второй член (штрафной) состоит из двух частей:

$$E_{constraint} = \sum_{ijkl} T_{ijkl}^{(1)} + T^{(2)}, \quad (4)$$

где

$$T_{ijkl}^{(1)} = \frac{\alpha}{2} [\delta_{ik}(1 - \delta_{jl}) + \delta_{jl}(1 - \delta_{ik})] = \frac{\alpha}{2} \left[\sum_{l \neq j} s_{ij} s_{il} + \sum_{k \neq i} s_{ij} s_{kj} \right], \quad (5)$$

$$T^{(2)} = \frac{\beta}{2} \left[\sum_{kl} s_{kl} - N \right]^2. \quad (6)$$

Первая часть штрафного члена устанавливает *запрет разветвлений* (т.е. запрет ситуаций, когда к одной точке присоединено больше одного сегмента-нейрона), а вторая *устанавливает баланс* между числом активных² нейронов и числом экспериментальных точек. Параметры α и β являются множителями Лагранжа³.

К сожалению, вышеприведенная схема Денби-Петерсона, принципиально запрещающая бифуркацию трека, неприменима в случае, когда мы имеем дело с рождением заряженных частиц в объеме детектора (от распада нейтральных каонов и гиперонов).

²Нейроны с величиной $s_{ij} > 0.1$.

³Обычно решения оказываются устойчивыми по отношению к этим параметрам [9].

Поэтому в ходе данной работы нам пришлось модифицировать штрафной член энергетической функции так, чтобы он позволял осуществлять распознавание разветвлений. Для соответствующего смягчения этой штрафной части в случае некоторого минимального количества пересечений, определяемого условиями эксперимента, мы использовали $T_{ijkl}^{(1)}$ только, *если количество активных нейронов N_{act} , имеющих общую точку с данным нейроном, превы-*

шало четыре, а в противном случае ($N_{act} < 4$) брали $\frac{T_{ijkl}^{(1)}}{2}$.

В соответствии с вышеупомянутой теоремой Хопфилда, для нахождения стационарного состояния сети требуется найти точку минимума энергетического функционала по значениям нейронов. Применение с этой целью метода градиентного спуска к энергетической функции (2) приводит к системе уравнений, определяющей динамику сети:

$$s_{ij} = \frac{1}{2} \left(1 + \text{sign} \left(-\frac{\partial E}{\partial s_{ij}} \right) \right). \quad (7)$$

Однако процедура итерационного решения системы (7) для случая бинарных нейронов часто приводит в какой-нибудь локальный минимум энергетической функции, что не является удовлетворительным решением, так как при этом значения многих нейронов, соответствующих ложным соединениям точек, все еще будут выше порога активности. Стандартным путем преодоления этой проблемы является введение статистического шума в систему с последующим применением теории среднего поля (см., например, [10]), что приводит к усреднению значений нейронов и замене ступенчатой функции (7) на функцию сигмоидного вида:

$$v_{ij} = \frac{1}{2} \left(1 + \tanh \left(-\frac{\partial E}{\partial v_{ij}} \frac{1}{T} \right) \right). \quad (8)$$

Здесь температура T соответствует уровню статистического шума. В соответствии с теорией среднего поля, нейроны $v_{ij} = \langle s_{ij} \rangle_T$ усредняются по температурному ансамблю, и значения дискретных нейронов становятся уже непрерывными в области $[0, 1]$.

Параметры α, β, T в выражениях (5), (6) и (8) являются подгоночными. С помощью системы уравнений (8) нейроны сети итеративно обновляются до достижения ею стабильного состояния. Однако то, что они теперь не являются бинарными, позволяет оперативно следить за теми из них, которые соединяют точки треков и под стимулирующим воздействием весов постепенно увеличивают уровень своей активности в процессе эволюции сети. В качестве порога уровня активности v_{ij} обычно выбирается $v_{min} = 0.1$.

Первоначально в данном исследовании была построена сеть, использующая пространственные координаты экспериментальных данных. Главное преимущество такой модели состоит в том, что вероятность совпадения и пересечения треков в пространстве намного меньше, чем вероятность совпадения или пересечения проекций на плоскости. К тому же при таком распознавании сразу получается пространственный трек и отпадает необходимость в сшивке проекций. Но из-за особенностей детектора число пространственных координат возрастает как квадрат числа треков, помноженный на число координатных камер. Такое количество пространственных точек получается при сопоставлении каждого отсчета одной сигнальной плоскости каждому отсчету другой сигнальной плоскости из той же пропорциональной камеры детектора⁴. К тому же все пространственные отсчеты расположены в пределах очень малого телесного угла (см. рис. 1). Все это вместе с довольно грубой дискретизацией (шаг между проволоками 2 мм), при которой близлежащие треки могут иметь общие точки, сильно затрудняет распознавание. Для сети появляется слишком много вариантов, и с энергетической точки зрения более выгодным может стать трек, проходящий по ложным отсчетам. Таким образом, вместо реального трека в качестве распознанного может быть взят близлежащий ложный трек. Число нейронов в сети равно $N(N - 1)$ (здесь N – количество экспериментальных точек, сегменты между которыми соответствуют нейронам), что дает для пяти треков на восьми координатных камерах сеть, состоящую из почти четырех тысяч нейронов. Такое количество нейронов тре-

⁴Каждая камера до магнита включает в себя две координатные плоскости.

бует больших затрат времени на стабилизацию состояния сети. Все вышесказанное делает применение пространственной модели сети неэффективным.

В итоге вместо одной сети, использующей пространственные координаты, была построена программная модель двух сетей, использующих результаты измерений отдельных плоскостей камер только по X- или только по Y-координате⁵. Такая модель имеет свои преимущества по сравнению с пространственной:

- Резко уменьшается число нейронов, а следовательно и время эволюции сети.
- Нет ложных отсчетов, возникающих в пространственной модели из-за сопоставления каждого отсчета по X каждому Y из той же камеры.

Построенная модель оказалась чувствительной к порядку обхода нейронов при вычислении новых значений v_{ij} для каждого нейрона. Наш опыт показал, что наиболее эффективным является следующий порядок вычислений:

- На первой итерации, когда штрафной член в энергетической функции (2) мал, рассчитываются показатели активности нейронов, предположительно относящихся к трекам. В данной работе таковыми считались нейроны, для которых E_{cost} в (2) имеет наибольшее значение. Это позволяет уже на второй итерации получить конфигурацию сети, достаточно близкую к минимуму, в которой нейроны, относящиеся к трекам, имеют значения, намного превосходящие значения "шумовых" нейронов.
- На последующих итерациях порядок обхода нейронов нужно изменить на обратный, так как штрафной член в энергетиче-

⁵Как принято в экспериментах ФВЭ с фиксированной мишенью, система координат выбрана так, что ось Oz направлена вдоль пучка перпендикулярно сигнальным плоскостям камер.

ской функции приобретает главную роль, и в случае последовательной эволюции нейронной сети наибольший штрафной балл получают нейроны, находящиеся в начале последовательности.

Итерации проводятся до тех пор, пока не будет выполняться

$$\frac{1}{N} \sum_{ij} \tilde{N} \left| \frac{v_{ij}(t+1) - v_{ij}(t)}{v_{ij}} \right| > \epsilon,$$

где \tilde{N} – число активных нейронов, т.е. нейронов, имеющих значения v_{ij} больше определенного порога, определяемого требованиями эксперимента. Итерационная процедура сходится при $\epsilon \sim 10^{-4}$.

В модели сегментов N экспериментальных точек приводят к $N(N-1)$ нейронам и такому же количеству уравнений. В принципе требуется N^3 операций для выполнения каждой итерации. Однако, вследствие локальной природы большинства задач поиска треков, это число может быть существенно уменьшено. Очень маловероятно, что две экспериментальные точки, находящиеся далеко друг от друга, будут соединены непосредственно. Поэтому можно ввести радиус обрезания R_{cut} , который характеризует область взаимодействия нейронов. Это приводит к уменьшению числа активных нейронов в 2 – 3 раза. При среднем числе активных нейронов \tilde{N} в области взаимодействия R_{cut} требуется $N\tilde{N}^2$ вычислений. Результаты не очень чувствительны к R_{cut} , но при его выборе надо учитывать ситуации, когда из-за неэффективности детектора в середине трека могут отсутствовать экспериментальные точки.

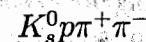
В конце эволюции активными считаются нейроны с $v_{ij} > v_{min} = 0.1$ (по ходу итераций для сохранения большего числа трек-кандидатов порог понижался и до $v_{min} = 0.05$). Затем конечное состояние подвергается процедуре "чистки", т.к. сеть может оставить некоторое число разветвлений.

Замечена чувствительность такой сети к наличию шумовых отсчетов. Для преодоления этого недостатка было предложено увеличивать вес тех нейронов-сегментов, вдоль которых есть допол-

нительные "подтверждающие" экспериментальные точки, в число которых должны входить отсчеты последних камер детектора, т.к. часть треков, появляющихся при распаде нейтральных частиц, может начинаться после первых камер. В результате удалось повысить робастность сети.

3 Результаты применения ИНС

Разработанный алгоритм распознавания прямолинейных треков при помощи ИНС и робастного метода восстановления их параметров был реализован на языке С. Такой выбор программной реализации был обусловлен требованием переносимости программного модуля на разные вычислительные платформы и возможностью встраивания данного модуля в программы, написанные на других языках программирования. Программа, реализующая алгоритм распознавания треков при помощи нейронной сети, была встроена в программный комплекс обработки эксперимента EXCHARM и протестирована на данных, смоделированных программой GEANT[11, 12]. При этом моделировался распад $\Lambda_c^+ \rightarrow$ бариона по каналу



с последующим распадом K_s^0 на два заряженных пиона.

Таким образом, при распаде Λ_c^+ образуется 5 заряженных частиц и, соответственно, должны регистрироваться 5 треков, три из которых проходят через всю экспериментальную установку, а два трека, соответствующие распаду K_s^0 на две заряженные частицы, начинаются в области распадного объема, включающего в себя область нескольких первых камер.

На рисунках 2 – 3 продемонстрированы результаты работы нейронной сети на моделированном событии с учетом неэффективности камер и с наложением шума. На всех рисунках приведено одно и то же событие, но в разном масштабе по вертикали

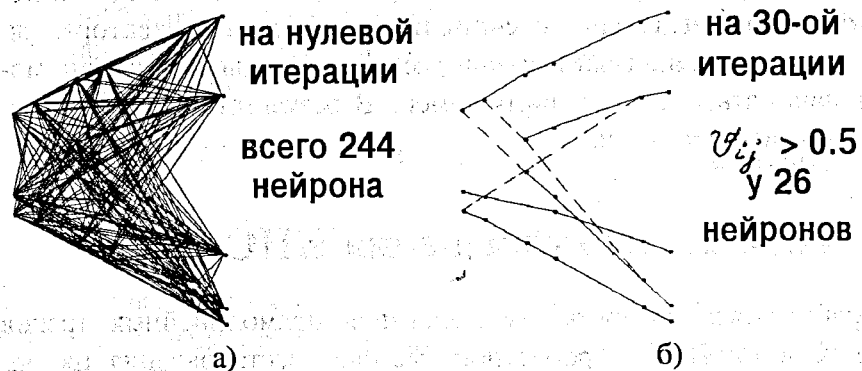


Рис. 2: Пример распознавания моделированного незашумленного события с Λ_c^+ -барионом в xOz -плоскости до магнита с помощью нейронной сети (масштаб А): а) исходное состояние сети; б) конечное состояние сети. Изображены только нейроны с показателем активности $v_{ij} > 0.5$.

для того, чтобы представить полную картину события. Реальные треки показаны сплошной линией, а ложные связи отмечены пунктиром.

Оценки эффективности распознавания проводились как на чистой модельной информации, так и с учетом реальных экспериментальных эффективностей пропорциональных камер, составляющей 80-90%, и при наложении шумов (3-5 фоновых точек на плоскость).

В нижеследующей таблице приведены результаты тестирования программы распознавания треков, использующей нейронную сеть, и программы массовой обработки эксперимента EXCHARM, использующей традиционный метод опорных дорожек до магнита. Проценты, приведенные в таблице, характеризуют количество полностью распознанных событий, т.е. тех событий, в которых были найдены все треки заряженных частиц до магнита.

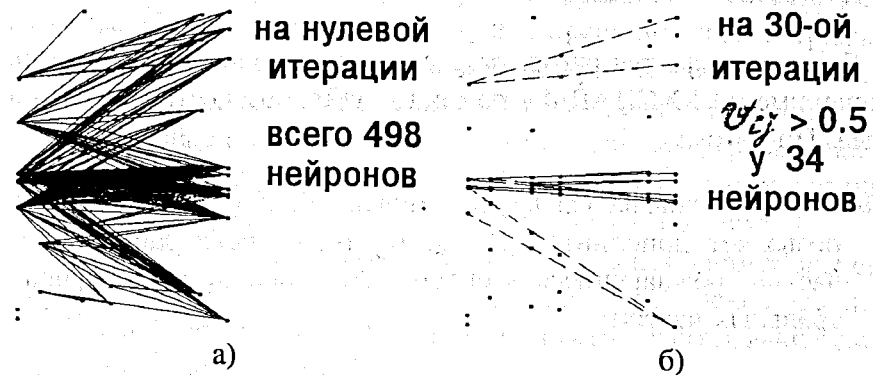


Рис. 3: Тот же пример, что и на рис.2, но с добавлением пяти шумовых точек на каждую плоскость (масштаб В): а) исходное состояние сети; б) конечное состояние сети.

Табл. Сравнительная эффективность распознавания событий.

Эффективность камер	Метод опорной дорожки	Нейронные сети	Комбинированный алгоритм
100%,	99%	99%	$\approx 100\%$
80 – 90%	98%	98%	$>99\%$
80 – 90% + фон	95%	97%	99%

В тяжелых фоновых условиях и экспериментальной неэффективности камер действующая программа обработки эксперимента EXCHARM не распознала 5.2% событий, а программа, основанная на ИНС, – 3.2%. При этом из нераспознанных событий для разных программ совпали только 0.8% от общего числа событий. Из этого результата следует вывод, что применение комбинированного алгоритма, использующего для распознавания большинства событий действующую программу обработки, а для распознавания сложных событий - программу ИНС, позволяет довести эффективность распознавания событий в реальных условиях до 99%.

4 Заключение

В работе предлагается алгоритм распознавания сложных событий в эксперименте EXCHARM с помощью нейронной сети типа Хопфилда. При этом:

- модифицирована сегментная модель Денби-Петерсона, что позволяет дополнительно распознавать треки заряженных частиц, рождающиеся в объеме детектора от распада нейтральных частиц;
- в соответствии с условиями эксперимента изменена эволюционная схема ИНС. После тщательного сравнительного исследования выбрана система из двух нейронных сетей в проекциях xOz и yOz при специальном порядке обхода нейронов, что обеспечило быструю сходимость эволюции ИНС, большую робастность и надежность реконструкции треков. Кроме того, в алгоритме ИНС введен локальный радиус обрезания, ограничивающий область взаимодействия нейронов, что значительно сократило число переборов на каждой итерации.

Результаты тестирования ИНС показали, что:

- при недостаточно хорошей работе камер детектора и в тяжелых фоновых условиях программа, использующая ИНС, имеет большую эффективность распознавания, чем действующая в эксперименте программа обработки, однако уступает ей в быстродействии;
- множества плохо распознанных событий для метода распознавания при помощи ИНС и традиционного метода опорных дорожек имеют малую область пересечения, что дает возможность построения эффективного комбинированного алгоритма восстановления треков при незначительном увеличении временных затрат на обработку.

Идея комбинированного алгоритма вкратце заключается в следующем. После распознавания x - и y -проекций треков традиционным методом опорных дорожек в программе обработки производится пространственная "сшивка" этих проекций. И если часть надежно найденных проекций не имеет сшивок, то включается нейросетевой алгоритм, который в большинстве случаев успешно решает задачу распознавания треков в таких событиях.

Алгоритм ИНС был опробован и на реальных данных эксперимента EXCHARM. Комбинированный алгоритм с использованием ИНС позволяет достичь 99% эффективности распознавания событий в реальных условиях.

Авторы выражают благодарность В.Д.Кекелидзе за поддержку и интерес к данной работе.

Работа выполнена при поддержке РФФИ, гранты 96-02-18734, 96-07-89190, 97-01-01027 и 98-07-90294.

Список литературы

- [1] А.Н.Алеев и др. (сотрудничество EXCHARM), *Наблюдение очарованного бариона Σ_c^0 в эксперименте ЭКСЧАРМ*, Краткие сообщения ОИЯИ, N 3, [77]-96, Дубна, 1986.
- [2] Н.Н.Говорун и др. *Программы распознавания прямолинейных треков, регистрируемых посредством искровых камер*, ОИЯИ, 10-7303, Дубна, 1973.
- [3] *Proc. of Int.Conf on Computing in High Energy Physics'92*, CERN 92-07, Geneve, 1992.
- [4] А.Ю.Бонюшкина, В.В.Иванов и др. *Идентификация событий со вторичной вершиной в эксперименте EXCHARM*, ОИЯИ, P1-96-56, Дубна, 1996.
- [5] И.М.Иванченко и др. *Математическое обеспечение для распознавания траекторий, регистрируемых многочастич-*

ным спектрометром БИС-2, ОИЯИ, P10-89-436, Дубна, 1989.

- [6] Hopfield J.J., *Learning Algorithms and Probability Distributions in Feedforward and Feed-back Networks*, Proc. Nat. Acad. Sci. SA, **84**, 8429 (1987).
- [7] C. Peterson, *Track finding with neural networks*. Nucl. Instr. and Meth. A279 (1986) 537.
- [8] B. Denby, *Neural networks and cellular automata in experimental high energy physics*. Comput. Phys. Commun., **49** (1988) 429.
- [9] C. Peterson and B. Söderberg, *A new method for mapping optimization problems onto neural networks*. Int. J. of Neural Systems, **1** (1989) 3.
- [10] И.В.Кисель, В.Н.Нескромный, Г.А.Осоков, *Применение нейронных сетей в экспериментальной физике*, ЭЧАЯ т.24,(1993), вып. 6, 1551.
- [11] R.Brun, F.Carminati, *GEANT Detector Description and Simulation Tool*, CERN Program Library Long Writeup W5013, September 1993.
- [12] Г.А.Аралбаева и др. *Комплекс программ, расширяющий возможности формализованного описания экспериментов в системе GEANT3*, ОИЯИ, P1-93-85, Дубна, 1985.

Рукопись поступила в издательский отдел
2 июля 1998 года.