

**сообщения  
Объединенного  
Института  
Ядерных  
Исследований  
Дубна**

**P10-87-243**

**Я.Балгансурэн, В.В.Глаголев, Н.Н.Говорун,  
А.Дирнер, В.Г.Иванов**

**ПУТИ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ РАБОТЫ  
СИСТЕМ ОБРАБОТКИ  
ФИЛЬМОВОЙ ИНФОРМАЦИИ**

**1987**

В течение ряда лет в Лаборатории вычислительной техники и автоматизации ОИЯИ ведутся работы по повышению уровня автоматизации процесса обработки фидьмовой информации<sup>/1/</sup>. Результатом этих работ явилось создание высокоавтоматизированных систем математической обработки результатов измерений камерных фотографий, позволяющих в 3-4 раза сократить календарные сроки анализа экспериментальных данных<sup>/2-4/</sup>. Это достигнуто в основном за счет оптимизации самого процесса обработки, алгоритмизации ряда операций, традиционно выполнявшихся людьми, и создания на этой основе ряда программ, внедрение которых позволило переложить на ЭВМ наиболее трудоемкие операции, выполнявшиеся ранее человеком<sup>/5-7/</sup>.

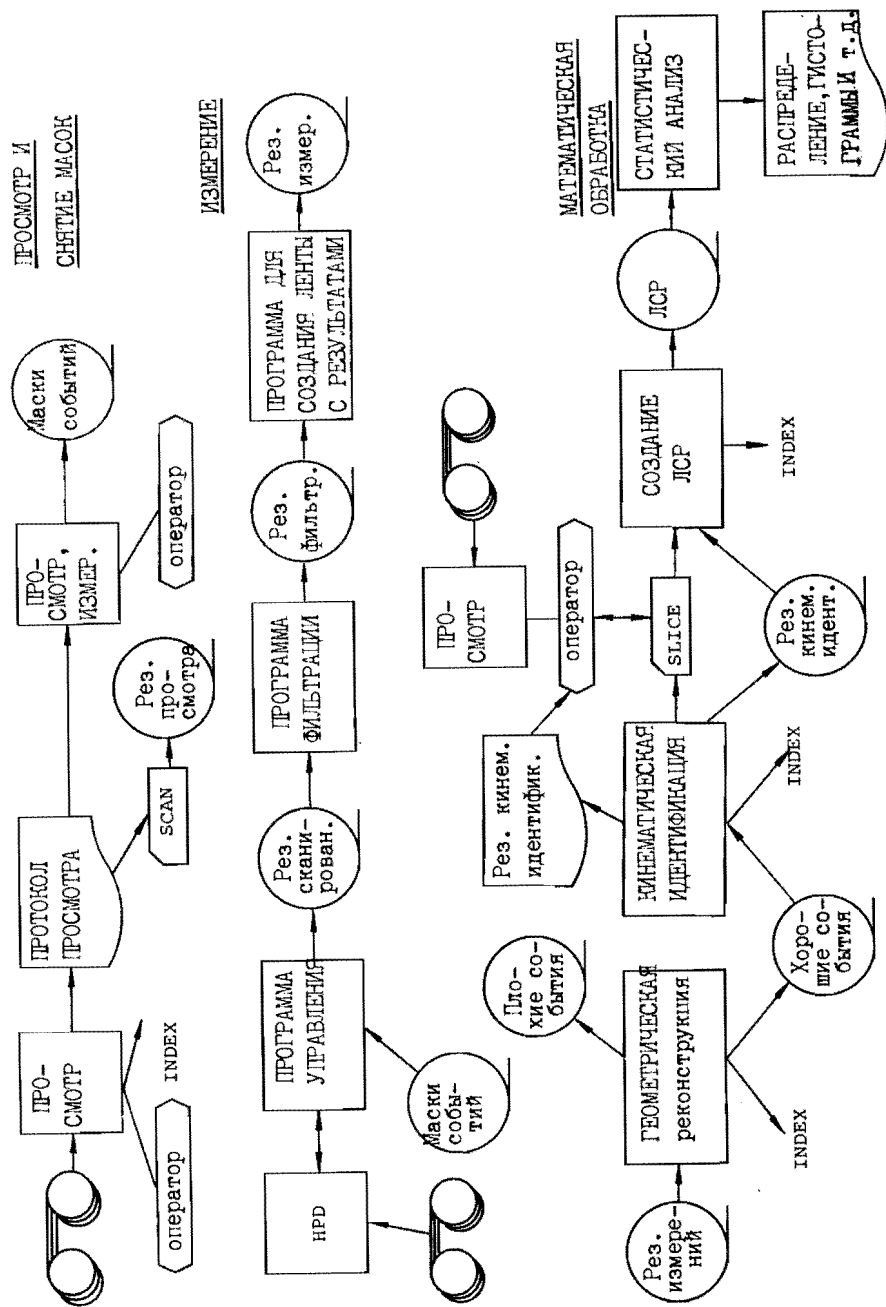
В данном сообщении обобщен опыт разработки такого типа систем, поскольку он может быть полезен при создании систем обработки не только фидьмовой информации, но и других экспериментальных данных.

#### § I. Традиционная схема процесса обработки фидьмовой информации и ее недостатки

Традиционная схема процесса обработки фидьмовой информации на базе сканирующего автомата типа НРД показана на рис. <sup>/8/</sup>. Схематично этот процесс можно разделить на следующие три этапа:

- I. Просмотр и снятие масок событий.
- II. Измерения элементов событий на стереоснимках.
- III. Математическая обработка результатов измерений, включая статистический анализ полученных данных.

I. Результатом работ первого этапа являются файл результатов просмотра и магнитные ленты, управляющие работой сканирующего автомата. В результаты просмотра для каждого найденного события заносятся следующие данные: номер фотопленки, номер стереокадра на фотопленке, число вторичных лучей и другие данные о топологии, ориентировочное положение его главной вершины и некоторая другая информация. Эти данные нужны для составления списков событий, которые нужно измерять. Кроме того, на заключительном этапе файл результатов просмотра используется для поиска потерянных в процессе обработки событий и проверки отсутствия систематического выброса событий каких-либо определенных типов.



П. Конечным результатом измерений являются координаты изображений элементов событий и опорных точек снимков, снабженные необходимой для их идентификации служебной информацией. Эти данные записываются на магнитные ленты и поступают на вход систем математической обработки<sup>8/</sup>.

Ш. В ходе математической обработки результатов измерений камерных фотографий восстанавливается пространственная картина измеренных событий и вычисляются параметры составляющих их треков для различных предположений о массах частиц. Затем для каждого события выделяется одна или несколько наиболее вероятных кинематических гипотез. Для этого сначала проверяется соответствие закону сохранения энергии-импульса параметров частиц для каждой из заданных для события гипотез. После этого производится дополнительный просмотр фотопленок, в ходе которого, наряду с кинематическими критериями, учитываются также и данные о плотностях почернения изображений треков. Информация об отобранных таким образом гипотезах переписывается на ленты суммарных результатов (ЛСР) эксперимента для последующего статистического анализа с помощью специальных карт выбора (Slice-карт).

Имеющиеся в ОИЯИ измерительные системы и мощные ЭВМ позволяют ежегодно обрабатывать сотни тысяч событий. Однако эти возможности используются далеко не полностью. Это объясняется тем, что в процессе обработки результатов измерений имеются этапы, требующие больших затрат ручного труда. К ним относятся разбор и анализ результатов счета на ЭВМ, переработка карт выбора, дополнительный просмотр, ведение журнала истории обработки событий. Дело усугубляется еще и тем, что значительная часть измеренных событий ( $\leq 30\%$ ) бракуется на различных этапах. Вследствие этого возникают новые задачи: анализ причин отказов, составление списков событий для повторных измерений, поиск и исключение из рассмотрения повторных измерений событий, успешно прошедших все этапы анализа.

Таким образом, традиционная схема обработки пленочной информации является многоступенчатой и требует больших затрат ручного труда. Использование систем программного ведения журнала истории обработки каждого найденного при просмотре события, хотя облегчает и ускоряет решение ряда задач, но тем не менее объем остающихся ручных работ все равно велик.

## § 2. Основные пути повышения эффективности работы систем обработки фильмовой информации

Традиционные системы обработки фильмовой информации являются большими и сложными человеко-машинными комплексами, работа которых в основном направляется и контролируется людьми. Однако последние не всегда успешно выполняют свои функции. Это, в основном, обусловлено следующими факторами:

I. Измерительные автоматы и ЭВМ "обрушивают" на людей такую лавину информации, что ее переработка нередко растягивается на годы и требует очень больших затрат ручного труда<sup>/9/</sup>. Так, например, на измерение 30 тыс. событий на сканирующем автомате типа nrd требуется около 15 недель, на обсчет результатов измерений - около 100 часов времени на ЭВМ CDC-6500, а на анализ получаемых результатов - 1,5-2,0 года. Это в значительной степени обусловлено большими объемами информации, которую нужно переработать. Так, например, печать результатов реконструкции и кинематической идентификации одного события занимает в среднем 2-3 страницы. Кроме того, на событие в среднем перфорировается 6-7 перфокарт. Поэтому при анализе только этих данных приходится просматривать десятки тысяч страниц машинной выдачи, перерабатывать сотни тысяч перфокарт.

II. Крайне низка производительность труда физиков при идентификации событий в ходе дополнительного просмотра. При напряженной работе физик за неделю идентифицирует 250-500 событий. Это в значительной степени связано с тем, что в ходе дополнительного просмотра приходится также просматривать и выдачу с ЭВМ, и отбирать гипотезы по физическим критериям (значения недостающих масс,  $X^2$  и т.п.).

III. Многоступенчатость процесса обработки и наличие многих операций, выполняемых людьми, требует от последних очень четкой работы, особенно при подготовке заданий ЭВМ<sup>/10/</sup>. Разного рода ошибки, допускаемые людьми, которые участвуют в работе такого рода комплексов, могут не только затягивать сроки обработки, но и приводить к существенным потерям ресурсов. Последние бывают значительными особенно тогда, когда ошибки обнаруживаются на заключительных стадиях анализа и его нужно повторять заново.

Особо следует остановиться на вопросе контроля получаемых результатов. В работе человеко-машинных комплексов по обработке фильмовой информации участвуют десятки людей разной квалификации. Поэтому здесь очень важен контроль получаемых данных на каждом из этапов. Он нужен для того, чтобы своевременно обнаруживать и устранять ошибки.

Для устранения отмеченных выше трудностей необходимо:

A. Снабдить человека такими средствами, которые позволят ему

оперативно перерабатывать поступающую с ЭВМ и автоматов информацию. Это можно сделать только с помощью ЭВМ, для чего нужно алгоритмизировать деятельность людей в процессе обработки фильмовой информации и создать на этой основе специальные программы, которые освободят людей от выполнения многих, если не всех, рутинных операций<sup>/11/</sup>.

B. Значительно ускорить процедуру идентификации событий. Это можно сделать за счет программного "просмотра" результатов кинематической идентификации событий, в ходе которого из имеющегося набора отбираются гипотезы, удовлетворяющие заданным критериям, включая данные об ионизации вторичных треков<sup>/5,12,13/</sup>.

B. Оптимизировать схему процесса обработки за счет сокращения числа выполняемых операций и уменьшения числа носителей информации. Сокращения числа выполняемых операций можно достичь путем объединения в одной программе функций, выполнявшихся ранее несколькими отдельными программами<sup>/14/</sup>, уменьшения числа сеансов счета вследствие увеличения числа событий, анализируемых в одном сеансе. Последнее требует резкого сокращения объема выдаваемой на печать информации и сроков анализа. Действительно, если сократить срок обработки какого-либо массива данных с нескольких месяцев до нескольких недель, то отпадает необходимость в анализе данных, получаемых на промежуточных этапах. Например, если идентификацию событий производить в течение 2-3 недель после получения результатов счета, то нет необходимости анализировать причины отказов сразу после кинематической идентификации, т.к. целесообразнее эту операцию производить после программного отбора физических гипотез с учетом ионизационных данных, т.е. практически перед формированием ЛСР, минуя промежуточные этапы. Крайне желательно исключить из процесса карты выбора физических гипотез, включив, скажем, данные о плотностях почернения треков событий в результаты просмотра, поскольку в нашем случае в процессе измерений это не делается.

Г. Перевести организацию обсчета данных по заданным цепочкам программ на машинную основу, т.е. ввести последовательную обработку данных под управлением ЭВМ, а не человека, сведя функции последнего по подготовке заданий для ЭВМ к минимуму.

Д. Организовать автоматический контроль получаемых на промежуточных этапах результатов с целью своевременного обнаружения ошибок и прекращения процесса анализа до их полного устранения. Этого можно достичь за счет последовательного внедрения в практику работы концепции программного сопровождения процесса обработки фильмовой информации<sup>/15/</sup>. Ее суть заключается в том, что после каждого из основных этапов процесса анализа специальная программа вычисляет средние значения ряда заданных величин и, сравнивая их с эталонными значениями, "решает" вопрос - вести анализ дальше или прекратить его.

Е. Обеспечить надежную работу программ обработки данных независимо от сбоев в работе ЭВМ, периферийного оборудования и ошибок операторов ЭВМ и лаборантов физических групп. Без решения этой задачи создать высокоавтоматизированную систему просто невозможно.

### § 3. Этапы работ по созданию высокоавтоматизированной системы математической обработки फिल्मовой информации

Первоначально характер взаимоотношений физиков и программистов в процессе создания систем обработки был таким. Программисты создавали или адаптировали программы для эксперимента, тестировали их на небольших группах событий и после этого передавали физикам. Последние вели по ним счет и после аварийных остановов обращались к программистам за помощью для выяснения причин и устранения обнаруженных ошибок. При этом довольно часто приходилось вносить в программы изменения, в значительной степени обусловленные ошибками в исходных данных, и процесс доводки программ занимал значительное время.

Проблема надежности работы программ стала особенно актуальной после перевода системы обработки на ЭВМ CDC-6500<sup>/16/</sup>. Это было связано с тем, что при массовой обработке данных требовалась надежная работа нескольких магнитофонов, поскольку исходные для счета данные считывались с магнитной ленты, а результаты счета записывались на одну или две магнитные ленты. Поскольку в наших условиях счет велся, как правило, небольшими порциями (100-300) событий, то перед каждым началом счета необходимо было устанавливать ленты в заданные позиции. Из-за не очень надежной работы магнитофонов и низкого качества имевшихся тогда магнитных лент эта операция достаточно часто срывалась. Поэтому наладить эффективную массовую обработку данных в наших условиях не удавалось. Считать же более крупными партиями было нельзя из-за больших объемов выдачи и довольно частых сбоев в работе программ из-за различного рода фатальных ошибок в результатах измерений. В связи с этим пришлось отказаться от работы с ленточными файлами и перейти на дисковые. Исходные данные с магнитной ленты переписывались на диск и на него же записывались результаты счета до тех пор, пока размеры файлов результатов не превышали заданных лимитов. После этого файлы с диска переписывались на ленты, диск чистился и процесс счета продолжался. Использование дисков позволило резко сократить потери времени ЭВМ, но несколько усложнило организацию обработки за счет появления новых операций (копирование ленты на диск, объединение файлов на диске, сброс файла с диска на ленту). Параллельно с этим велись работы по сокращению аварийных остановов программ за счет расширения их диагностических возможностей и перехвата фатальных ошибок. Все это позволило наладить достаточно

надежную работу программ, но усложнило и без того достаточно сложную систему. Вследствие этого возросли потери времени ЭВМ за счет участвовавших ошибок людей, готовивших ей задания.

Поэтому встал вопрос о создании системы, позволяющей организовать последовательный обсчет данных по заданной цепочке программ под управлением ЭВМ, начиная с копирования лент результатов измерений на диск и кончая формированием ЛСР. Первоначально планировалось, что система будет надстройкой над имеющимися программами, не затрагивая последних. Что касается информации, необходимой для продолжения счета, то ее предполагалось передавать от программы к программе или одного сеанса счета к другому через специально создаваемые временные файлы. Последовательность вызова программ задавалась графом выполняемых операций. Пользователю при обращении к системе необходимо было лишь указать выделенный задаче ресурс времени и, если требовалось, номера магнитных лент<sup>/17/</sup>. Система успешно прошла опытную эксплуатацию силами разработчиков, но не оправдала себя в ходе массового счета. Оказалось, что надежность работы CDC-6500 была недостаточна для успешного функционирования такой системы. Наиболее слабым ее звеном оказался временный файл для передачи информации от одной программы к другой или одного сеанса счета к другому. Из-за зависаний системы случалось, что содержащаяся в нем информация не соответствовала реальной ситуации. Иначе говоря, система оказалась не защищенной от сбоев в работе ЭВМ и ошибок операторов. В связи с этим были проведены следующие работы:

И. Повышена надежность работы основных программ системы по отношению к внешним сбоям, включая ошибки операторов. Это было достигнуто за счет включения в эти программы специальных подпрограмм, предназначенных для нахождения файлов исходных данных и результатов счета, установки их в нужные позиции и организации автоматического продолжения файлов результатов. Т.е. программы сами стали "находить", где брать данные, сколько записей пропускать. Вследствие этого временный файл для передачи служебной информации от одной программы к другой стал ненужен.

П. Процесс математической обработки был разделен на несколько последовательных этапов, которым соответствуют отдельные цепочки программ. Передача управления от одной программы к другой на ЭВМ CDC-6500 стала производиться с помощью управляющих процедур языка ССЛ, посредством которых и осуществляется общение пользователя с системой. С помощью небольшого набора управляющих процедур удалось наладить обсчет данных в режиме программного управления и существенно упростить работу пользователей по подготовке заданий для ЭВМ. Это позволило наладить практически безаварийную работу системы с коэффи-

циентом использования полезного времени ЭВМ, близким к 100%. Таким путем были решены задачи организации счета под управлением ЭВМ и обеспечения надежной работы программ системы.

Параллельно велись работы по алгоритмизации деятельности человека и созданию на этой основе специальных программ, позволяющих заменить его труд при решении многих рутинных задач<sup>/11/</sup>.

Первым шагом на этом пути явилась разработка и реализация методики программного отбора физических гипотез<sup>/12/</sup>, являвшейся тогда наиболее узким местом системы, о чем уже говорилось в предыдущем параграфе. Для ускорения и упрощения этой процедуры была создана специальная программа, решающая следующие задачи:

- отбор физических гипотез, удовлетворяющих заданным критериям;
- упорядочение физических гипотез для неоднозначно идентифицированных событий по уменьшению вероятностей их наблюдения;
- выдача информации об оставшихся физических гипотезах в удобной для идентификации событий форме.

Внедрение этой программы в практику работы позволило резко сократить выдачу на печать и перфокарты и в 3-4 раза ускорить процесс идентификации событий. Хотя это значительно ускорило работу, осталось еще одно узкое место, связанное с использованием для выбора физических гипотез на ЛСР карт отбора. Для их исключения из процесса анализа было решено производить визуальную оценку плотностей почернения вторичных треков событий не при дополнительном, а при предварительном просмотре и заносить соответствующие данные в файл результатов просмотра. Это позволило осуществить отбор физических гипотез не только по кинематическим, но и ионизационным критериям, без карт отбора, сократить объем выдаваемой на печать информации, в несколько раз уменьшить число событий, нуждающихся в дополнительном просмотре<sup>/5/</sup>, а также автоматизировать процедуру формирования ЛСР<sup>/7/</sup>.

Достаточно много времени и сил в процессе обработки данных в камерных экспериментах уходит на подведение итогов обработки как отдельных массивов, так и всего эксперимента в целом<sup>/6/</sup>. Под этим понимается анализ полученных результатов и составление различного рода списков. Таких, например, как хорошо измеренных событий; забракованных, которые нужно перемерять; плохих, которые нужно просто исключить из дальнейшего рассмотрения; два или более раз измеренных хороших событий (дубли). Кроме того, на заключительном этапе нужно получить список потерянных в процессе обработки событий. Для того чтобы автоматизировать все операции, связанные с составлением списков, было сделано следующее:

- по соглашению с физиками к плохим (неизмеряемым) стали относить те события, которые дважды браковались в ходе анализа по одной и той же причине;

- события, забракованные один или два раза, но по разным причинам, включались в список для перемеров;

- на файл результатов любой программы, кроме последней (SLICE), записывалась информация о всех поступивших на вход системы событиях. Причем для забракованных - только необходимая для их идентификации служебная информация и код обнаруженной ошибки. Это позволило с помощью одной программы PASPORT получать интересующие пользователей списки после любого этапа заданного массива<sup>/6/</sup>;

- для поиска и исключения дублей, а также составления различного рода списков на заключительном этапе обработки создается специальный журнал истории анализа всех поступивших на вход системы событий. Сопоставляя данные этого журнала и файла результатов просмотра, специальная программа находит и составляет список событий, которые были найдены при просмотре, но не попали на вход системы;

- для проверки получаемых в ходе анализа данных была разработана программа TEST, с помощью которой гистограммируются и вычисляются средние значения ряда физических величин, характеризующих тот или иной процесс ( $\chi^2$ , мм<sup>2</sup> и т.п.)<sup>/18/</sup>.

В настоящее время завершены работы по созданию версии программы обработки результатов просмотра в режиме диалога человека с ЭВМ.

#### Заключение

О тех возможностях, которые предоставляют рассмотренные в работе системы обработки данных, широко использующие возможности современных ЭВМ для решения различного рода задач, можно судить по следующему примеру.

В течение полугода было проведено облучение пузырьковой камеры в пучке поляризованных дейтронов, проведены просмотр и измерения 20 тыс. событий, включая перемеры, сформирована ЛСР и произведен анализ полученных результатов. На основе последней была получена оценка векторной поляризации дейтронов  $P_2 = 0,50 \pm 0,06$ , а также отработана методика развития эксперимента.

Разработанная система достаточно устойчива по отношению к различного рода внешним помехам и в дальнейшем будет усовершенствоваться по мере поступления дополнительных технических средств.

#### Литература

1. Говорун Н.Н., Иванов В.Г., Позе Р. В кн.: Научное сотрудничество социалистических стран в ядерной физике. Москва, Энергоатомиздат, 1986, с. II2-II9.

2. Балгансурэн Я. и др. ОИЯИ, Р10-85-516, Дубна, 1985.
3. Абдурахимов А.У. и др. ОИЯИ, Р10-86-612, Дубна, 1986.
4. Балгансурэн Я. и др. ОИЯИ, Р10-86-706, Дубна, 1986.
5. Балгансурэн Я. и др. ОИЯИ, Р10-86-412, Дубна, 1986.
6. Балгансурэн Я. и др. ОИЯИ, Р10-86-804, Дубна, 1986.
7. Балгансурэн Я. и др. ОИЯИ, Р10-86-805, Дубна, 1986.
8. Villemoes P. In: Proc. of 1970 CERN Computing and Data Processing School. Geneva, CERN, 1970, p. 105-206.
9. Бадалян С.Г. ОИЯИ, Р10-12879, Дубна, 1979.
10. Глаголев В.В. и др. ОИЯИ, Р10-82-460, Дубна, 1982.
11. Говорун Н.Н., Иванов В.Г. В кн.: Обработка физической информации. Тезисы докладов III всесоюзного семинара по обработке физической информации. Цахкадзор. ЦНИИатоминформ - Ереван, 1985, с.3-6.
12. Глаголев В.В. и др. ОИЯИ, Р10-82-369, Дубна, 1982.
13. Балгансурэн Я. и др. ОИЯИ, Р10-85-510, Дубна, 1985.
14. Бадалян С.Г. и др. ОИЯИ, Р10-12474, Дубна, 1979.
15. Балгансурэн Я. и др. В сб.: Труды V международного совещания по проблемам математического моделирования, программированию и математическим методам решения физических задач. Д10, II-84-818, Дубна: ОИЯИ, 1984, с.330-332.
16. Бадалян С.Г. и др. ОИЯИ, Р10-11448, Дубна, 1978.
17. Глаголев В.В. и др. ОИЯИ, Р10-82-211, Дубна, 1982.
18. Бадалян С.Г. и др. ОИЯИ, Р10-12606, Дубна, 1979.

## НЕТ ЛИ ПРОБЕЛОВ В ВАШЕЙ БИБЛИОТЕКЕ?

Вы можете получить по почте перечисленные ниже книги, если они не были заказаны ранее.

D9-82-664	Труды совещания по коллективным методам ускорения. Дубна, 1982.	3 р. 30 к.
D3,4-82-704	Труды IV Международной школы по нейтронной физике. Дубна, 1982.	5 р. 00 к.
D11-83-511	Труды совещания по системам и методам аналитических вычислений на ЭВМ и их применению в теоретической физике. Дубна, 1982.	2 р. 50 к.
D7-83-644	Труды Международной школы-семинара по физике тяжелых ионов. Алушта, 1983.	6 р. 55 к.
D2,13-83-689	Труды рабочего совещания по проблемам излучения и детектирования гравитационных волн. Дубна, 1983.	2 р. 00 к.
D13-84-63	Труды XI Международного симпозиума по ядерной электронике. Братислава, Чехословакия, 1983.	4 р. 50 к.
D2-84-366	Труды 7 Международного совещания по проблемам квантовой теории поля. Алушта, 1984.	4 р. 30 к.
D1,2-84-599	Труды VII Международного семинара по проблемам физики высоких энергий. Дубна, 1984.	5 р. 50 к.
D17-84-850	Труды III Международного симпозиума по избранным проблемам статистической механики. Дубна, 1984. /2 тома/	7 р. 75 к.
D10,11-84-818	Труды V Международного совещания по проблемам математического моделирования, программированию и математическим методам решения физических задач. Дубна, 1983.	3 р. 50 к.
	Труды IX Всесоюзного совещания по ускорителям заряженных частиц. Дубна, 1984 /2 тома/	13 р. 50 к.
D4-85-851	Труды Международной школы по структуре ядра, Алушта, 1985.	3 р. 75 к.
D11-85-791	Труды Международного совещания по аналитическим вычислениям на ЭВМ и их применению в теоретической физике. Дубна, 1985.	4 р.
D13-85-793	Труды X Международного симпозиума по ядерной электронике. Дубна, 1985.	4 р. 80 к.
D3,4,17-86-747	Труды V Международной школы по нейтронной физике. Алушта, 1986.	4 р. 50 к.

Заказы на упомянутые книги могут быть направлены по адресу:  
101000 Москва, Главпочтамт, п/я 79  
Издательский отдел Объединенного института ядерных исследований

Рукопись поступила в издательский отдел  
14 апреля 1987 года