

сообщения
объединенного
института
ядерных
исследований
Дубна

77/
2-82

4/1-82

10-81-699

В.В.Глаголев, Н.Н.Говорун, В.Г.Иванов,
Т.А.Стриж

ОРГАНИЗАЦИЯ КАТАЛОГА
ОБРАБАТЫВАЕМЫХ СОБЫТИЙ
НА БАЗЕ ФАЙЛОВ ПРЯМОГО ДОСТУПА

1981

Для оперативного получения информации о текущем состоянии дел в системе обработки фильмовой информации необходимо фиксировать историю каждого найденного при просмотре события. Здесь под историей события понимаются результаты его прохождения через различные этапы процесса обработки, начиная с просмотра и кончая записью на ленту суммарных результатов.

Поскольку число обрабатываемых в камерных экспериментах событий обычно составляет десятки, а то и сотни тысяч, широкое распространение получили системы типа INDEX ^{1/}, позволяющие вести "журналы" с историей обработки, организованные в виде файлов последовательного доступа на магнитных лентах или дисках ^{12,3/}.

Специфика организации систем математической обработки фильмовой информации с жидководородных пузырьковых камер, связанная с широким использованием частных дисковых пакетов ^{17/}, потребовала разработки специального варианта системы типа INDEX на базе файлов прямого доступа, которая была реализована на ЭВМ CDC-6500. Данная работа посвящена вопросам организации этой системы, которая используется в настоящее время в экспериментах, проводимых на жидководородных пузырьковых камерах ОИЯИ.

§ I. Организация обработки фильмовой информации в рамках системы INDEX

На рис. I показана организация обработки фильмовой информации на базе варианта системы INDEX, которая была разработана в ЦЕРНе ^{14/}.

Результаты просмотра (номера событий, топология, положение главных вершин и ряд других данных) записываются на магнитную ленту IST (Index Scan Tape).

Результаты измерений фотоплёнок, производимых на автоматических или полуавтоматических устройствах, после предварительной обработки

записываются на магнитные ленты MRT (Measurement Result Tape) в форме, непосредственно пригодной для пространственной реконструкции. Обсчет данных, содержащихся на MRT, производится по хорошо известной цепочке программ: THRESH-GRIND-SLICE.

При работе в системе INDEX программы GRIND и SLICE имеют по две ленты результатов. На магнитные ленты GLT (Grind Library Tape) и DST (Data Summary Tape) записываются основные результаты счета по программам GRIND и SLICE соответственно. История же событий фиксируется на лентах GUI (Grind Update Index) и SUI (Slice Update Index), на которые записываются банки со специальной информацией.

В кинематический банк события программа GRIND заносит следующие данные:

- номер эксперимента;
- номер события;
- дату и время счета;
- X- и Y-ые координаты главной вершины;
- порядковый номер события на ленте результатов (GLT);
- номер ошибки, если событие было забраковано при реконструкции или идентификации.

В банк послекинематических результатов помимо номера эксперимента и номера события программа SLICE записывает следующее:

- дату и время счета по программе GRIND;
- дату счета по программе SLICE.

Кроме того, в в любой из указанных выше банков можно записывать дополнительную текстовую информацию.

На основе данных, содержащихся на магнитных лентах IST, GUI, SUI, программа INDEX создает новую ленту MIT (Master Index Tape), на которой содержится информация о каждом обрабатываемом событии. Эта информация хранится в банках четырех типов. Сюда относятся:

- главный банк события, в который записывается его номер и количество банков каждого из остальных типов;
- банки результатов просмотра;
- банки кинематических результатов;
- банки послекинематических результатов.

Информация на MIT упорядочивается в порядке возрастания номеров событий.

Примечание: В системе имеется возможность создавать банки кинематических результатов с GLT.

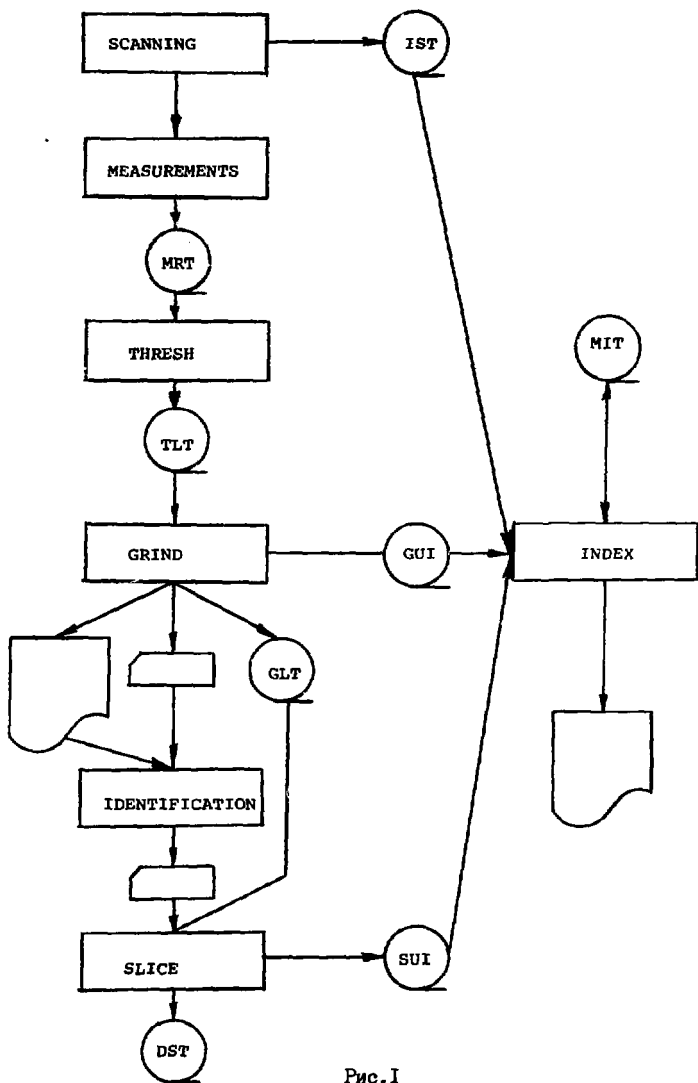


Рис. I

Таким образом, MIT является как бы "физическим журналом", в котором постепенно накапливается информация об истории обработки событий.

Главной задачей программы INDEX является создание MIT и обновление содержащейся на ней информации по мере поступления. INDEX позволяет добавлять, вычеркивать, изменять, расширять и объединять информацию во всех банках этой ленты, а также выдавать на печать содержащуюся на ней информацию по запросам экспериментаторов.

Недостатком системы такого типа с точки зрения ее эксплуатации является то, что для работы требуется несколько магнитных лент, а такие задачи на современных ЭВМ имеют, как правило, достаточно низкий приоритет. Кроме того, для обеспечения сохранности получаемых данных необходимо изготавливать копии MIT после записи на нее новой порции информации. Вследствие этого вариант системы INDEX созданный в ОИЯИ на ЭВМ БЭСМ-6, не нашел широкого применения ¹⁵⁷.

Переход к использованию дисковых файлов вместо файлов на магнитных лентах улучшает ситуацию, но требует достаточно больших ресурсов памяти, т.к. информация о каждом событии занимает на MIT несколько десятков слов.

Поэтому с учетом специфики организации обсчета данных на ЭВМ CDC-6500 и на базе имеющихся программных и технических средств была разработана система STATUS, обладающая большинством возможностей системы INDEX, но реализованная на совершенно других принципах.

§ 2. Организация математической обработки фильмовой информации с жидководородных пузырьковых камер ОИЯИ на ЭВМ CDC-6500

В настоящее время математическая обработка фильмовой информации с жидководородных пузырьковых камер ОИЯИ в основном производится на ЭВМ CDC-6500.

Для физических групп, ведущих обработку событий с камер ВК-100 и "Лидмила", выделено по частному дисковому пакету емкостью около 10 млн. слов. На дисках хранятся библиотеки программ, по которым ведется массовый обсчет данных, и на них же записываются исходные данные и промежуточные результаты счета. Кроме того, несколько процентов от общего ресурса памяти можно использовать для ведения журнала, который в дальнейшем будем называть каталогом обрабатываемых событий (КОС).

Упрощенная схема организации процесса обработки фильмовой информации с однометровой жидководородной пузырьковой камерой ОИЯИ показана на рис. 2 ¹⁷¹.

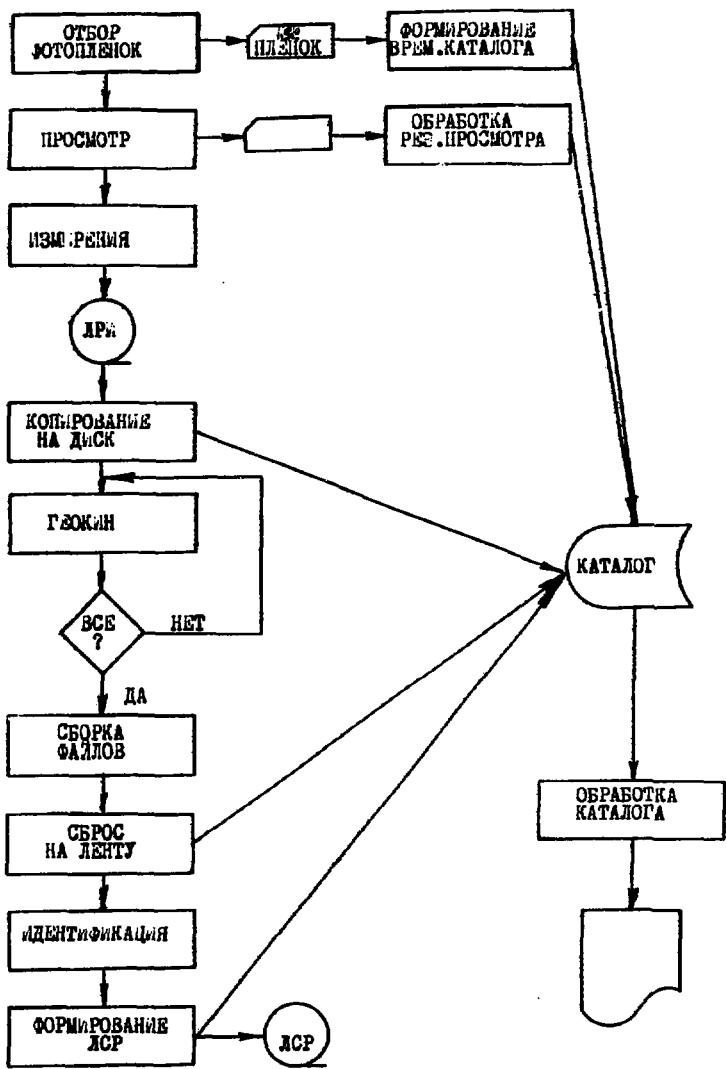


Рис. 2

Прежде всего определяются число и номера фотопленок, которые будут обрабатываться в данном эксперименте, а также устанавливается максимальное число событий, которые могут быть зафиксированы на одной фотопленке. На основании этих данных формируется временный каталог, в котором для каждой фотопленки отведено постоянное число слов, рассчитанное на максимально возможное число событий.

В ходе просмотра фотопленок находятся стереоснимки, на которых имеются изображения событий исследуемых тел. Для последующей идентификации каждому событию присваивается номер, в который входит номер рулона фотопленки, номер стереоснимка на пленке и порядковый номер события на снимке. Кроме того, для каждого события фиксируется топология, зона, в которой находится главная вершина, прибор (НРД или ПУОС), на котором его следует измерять, и номер просмотра. Часть найденных событий не может быть измерена из-за встречающихся на фотопленках дефектов (пятна, царапины, разрывы). Эти события включаются в общий список, но метятся специальным символом (BAD). Таким образом, найденные при просмотре события делятся на следующие три группы:

- измеряемые на НРД;
- измеряемые на полуавтоматах;
- неизмеряемые.

Принадлежность события к каждой из этих групп фиксируется в его истории, т.к. это позволяет составлять списки событий для измерений и контролировать их поступление с помощью ЭВМ.

Найденные при просмотре события измеряются на сканирующем автомате типа НРД или полуавтоматах. Результаты измерений или сканирования, обработанные по соответствующим программам, записываются на ленты результатов измерений (ЛРи) в формате исходных данных для программы ГЕОКИН.

Содержимое ЛРи затем копируется на магнитный диск, а в каталоге фиксируются все поступившие после измерений события.

Восстановление пространственной картины событий и их кинематическая идентификация производится, как правило, с помощью программы ГЕОКИН, которая считывает с диска исходные данные и на него же записывает результаты счета. Всего на одной ЛРи может находиться от одной до полутора тысяч событий, на обсчет которых требуется несколько часов времени на ЭВМ CDC-6500. Поскольку обсчет событий производится небольшими партиями, по несколько сот в каждой, то результаты счета накапливаются на магнитном диске в виде отдельных файлов, затем объединяются и сбрасываются на магнитную ленту. На этом этапе

обработки в каталоге фиксируется принадлежность события к одной из следующих групп:

- хорошие события; к ним относятся события, в которых после кинематической идентификации осталась одна или несколько гипотез.
- плохие события; к этой группе программа ГЕОКИН относит события, в измерениях которых были обнаружены какие-либо ошибки или найдено несоответствие измеренных величин, например параметров пучкового трека, заданным значениям. Результаты счета и стереоснимки плохих событий просматриваются физиками, которые либо отправляют их на повторные измерения, либо переводят в группу неизмеримых.

Таким образом, после сброса на ленту результатов счета по программе ГЕОКИН в каталоге зафиксирована принадлежность события к хорошим или плохим. Кроме того, для последних в каталоге также фиксируется код обнаруженной ошибки.

Заключительный этап обработки - формирование лент суммарных результатов, на которые переписываются отобранные физиками гипотезы о каждом рассматривавшемся событии. Эта задача решается в следующей последовательности. Сначала производится дополнительный просмотр фотопленок, в ходе которого сравнивается наблюдаемая на снимках ионизация с вычисленной для заданных наборов масс (гипотез). Эта процедура получила название идентификации. Затем с помощью специальной программы отобранные гипотезы переписываются на ЛСР. В ходе дополнительного просмотра также выявляются события, которые следует исключить из рассмотрения или направить на повторные измерения. Если в каталоге зафиксировано, что событие переписано на ЛСР, то процесс его обработки завершен. Таким образом, после завершения процесса обработки всего массива событий в каталоге должны находиться события только двух типов: переписанные на ЛСР и неизмеримые. Если же их сумма не равна общему числу найденных при просмотре событий, то часть из них потерялась на каком-либо этапе и следует выяснить причину потерь.

§ 3. Назначение и принципы организации каталога обрабатываемых событий

Основное назначение каталога - фиксировать результат прохождения каждого найденного при просмотре события через следующие этапы процесса обработки /7/:

- измерения;
- расчет по программе ГЕОКИН;
- запись на ЛСР.

На каждое событие в каталоге отводится три слова. Текущее состояние события фиксируется с помощью определенных битов его статусного слова. Биты этого слова делятся на две группы: статусные и информационные. Назначение первых — фиксировать статус события (измеримое или неизмеримое, поступило с измерительной установки или нет, прошло через ГЕОКИН или забраковано и направлено на перемер). Информационные биты предназначены для более подробного описания истории события (номер просмотра, в котором оно было найдено, название измерительного прибора, тип ленты суммарных результатов и т.п.). Назначение битов статусного слова для α_p -эксперимента указано в таблице I.

Для α_p -эксперимента, проводимого на камере "Лидыла", значения битов в основном те же, за исключением 3, 57 и 58, которые отведены для регистрации событий, имеющих останавливающиеся треки и стриппинговые частицы.

Введение в каталог статусных слов существенно сократило его размеры по сравнению с первоначальным вариантом системы INDEX и позволяет хранить его на том же дисковом пакете, который используется для хранения исходных данных и промежуточных результатов счета.

Для идентификации событий в каталоге используются два слова. В первое записывается его номер и топология, которая вычисляется по следующей формуле:

$$1000 \times NT^2 + 100 \times NSI + 10 \times NV^0 + NV^1$$

Здесь: NT^2 — число лучей первичной звезды,
 NSI — число вторичных взаимодействий,
 NV^0 — число v^0 -частиц,
 NV^1 — число изломов.

Во второе — число лучей первичной звезды, не считая пучкового, метки и координаты главной вершины в пространстве или на первом стереоснимке.

Сам каталог организован в виде файла, обеспечивающего прямой доступ к информации, которая относится к одной фотопленке.

§ 4. Критерии для отождествления событий в процессе обновления каталога

Как уже отмечалось, каждому найденному при просмотре событию присваивается номер, используемый затем в качестве его идентификатора. Этот номер составляется из номера фотопленки, снимка и порядкового номера события на снимке. Если бы порядковый номер события на

Таблица I

№ разряда	Значение
I	Найдено в первом просмотре
2	Найдено во втором просмотре
3	Найдено в третьем просмотре
4	В событии имеется χ_{me} -распад
5	Измерено на НРД
6	Измерено на полуавтомате
7	Измерено на СИ
8	Измерено один раз
9	Измерено два раза
10	Измерено три раза
11	Результат ГЕОКИНа - FIT -гипотеза
12	Результат ГЕОКИНа - NOFIT -гипотеза
13	Забраковано ГЕОКИН-ом
14	Неизмеримое
15	Записано на ЛСР фитированных гипотез
16	Записано на ЛСР нефитированных гипотез
17	Обсчитано на ЭВМ CDC-6500
18	Образовано непучковым треком
19	Обсчитано на ЕС-1040
20	Прошло стадию просмотра
21	Прошло стадию измерений
22	Прошло через ГЕОКИН (не забраковано)
23	Направлено на повторные измерения
24	Переписано на ЛСР
25	В событии имеется пара Далица
26	В событии есть вторичное взаимодействие
27	В событии есть ν^0 -частицы

Таблица I (продолжение)

№ разряда	Значение
28	В событии имеются γ -кванты
29	В событии есть нейтральное взаимодействие
30	В событии есть излом
3I-40	код ошибки для первого измерения
4I-50	код ошибки для второго измерения
5I-57	код ошибки для третьего измерения

снимке сохранился в процессе измерений, то все события можно было бы идентифицировать по их номерам. Однако, как вскоре выяснилось, в нашем случае довольно часто после измерений разные события одного стереоснимка имеют одинаковые номера. Поэтому для αp -эксперимента была принята следующая процедура маркировки главных вершин событий. Главная вершина первого по порядку события метится буквой А, второго - В, третьего - С и четвертого - D. В тех случаях, когда соотношение между меткой вершины и ее порядковым номером не выполняется, вершине присваивается номер, соответствующий метке. Однако при перемерах событий им всегда присваивается порядковый номер, равный единице, и соответствующая ему метка - независимо от того номера и метки, которые оно имело первоначально. Поэтому когда на снимке имеется несколько событий и первое из них успешно прошло программу ГЕОКИН, то перемеры остальных будут считаться дублями первого и игнорироваться при обновлении каталога, если при идентификации использовать только номер события и метку главной вершины. В случае, если программа обновления каталога при занесении в него данных об измеренных событиях обнаруживает, что событие с данным номером успешно прошло через программу ГЕОКИН, проверяется, нет ли на этом стереоснимке других событий. Если таковые имеются и в каталоге зафиксировано, что они были посланы на перемер и еще не обчислены по программе ГЕОКИН, то поступившее событие отождествляется с первым по порядку событием этого снимка, находящимся на стадии измерений или перемеров и имеющим ту же топологию. Если таких событий в каталоге нет, или нет признака перемера, или нет совпадения топологий, результаты измерений игнорируются.

Такова в общих чертах процедура отождествления событий по результатам измерений. Следует отметить, что работа в системе STATUS требует строгого выполнения правил маркировки событий.

После счета по программе ГЕОКИН отождествление производится по тем же критериям, что и на предыдущем этапе. Для отождествления событий, направляемых на перемер, необходимо, чтобы в программе ГЕОКИН приводились в соответствие порядковый номер события и метка его главной вершины, т.к. в результатах счета для плохих событий указываются их номера и коды ошибок.

При отождествлении событий, записанных на ленты суммарных результатов, помимо номеров событий и топологий также используются значения пространственных координат главных вершин. Поэтому на этом этапе процедура отождествления не представляет затруднений.

Заключение

Приведем в заключение некоторые данные, характеризующие требующийся для работы системы ресурс памяти, а также затраты времени ЭВМ на выполнение различных операций.

Так, например, постоянный каталог для 28 тыс. событий занимает около 80 тыс. слов.

В таблице 2 приведены данные о затратах времени центрального процессора ЭВМ СРС-6500 на обновление содержимого каталога на различных этапах процесса обработки.

Таблица 2

Тип информации	Число обработ. событий	Время в сек.
Измерения	1904	64
ГЕОКИН	3918	52
Перемеры	2072	43
ЛСР	6788	99

Рассмотренная в работе система ведения каталога в течение нескольких лет используется в экспериментах, проводимых на однометровой жидководородной камере /8/, закончена также ее адаптация для $\bar{\alpha}$ p - эксперимента, проводимого на камере "Людмила".

Литература

1. Letertre C. В кн. CERN COMPUTER 6000 SERIES PROGRAM LIBRARY. PROGRAM INDEX, CERN, Geneva, 1970.
2. Позе Р.А. В кн.: Материалы второго всесоюзного семинара по обработке физической информации. ЕРФИ, Ереван, 1978, с.191.
3. Горичев П.А. и др. Препринт ИФВЭ, СЛК 72-II7, Серпухов, 1972.
4. Villemoes P. В кн.: Proceedings of the 1970 CERN Computing and Data Processing School. CERN, Geneva, 1970, с. 105.
5. Буздавина Н.А. и др. В кн.: Материалы семинара по обработке физической информации. ЕРФИ, Ереван, 1976, с.168.
6. Бадалян С.Г. и др. ОИЯИ, IO-II448, Дубна, 1978.
7. Абдурахимов А.У. и др. ОИЯИ, IO-8I-359, Дубна, 1981.
8. Буздавина Н.А. и др. ОИЯИ, I-8I-530, Дубна, 1981.

Рукопись поступила в издательский отдел
9 ноября 1981 года.