

СООБЩЕНИЯ  
ОБЪЕДИНЕННОГО  
ИНСТИТУТА  
ЯДЕРНЫХ  
ИССЛЕДОВАНИЙ

Дубна

Р11-99-252

В.М.Котов, Н.М.Никитюк

ПРИМЕНЕНИЕ АТМ-ТЕХНОЛОГИИ  
В ЭКСПЕРИМЕНТАЛЬНОЙ ФИЗИКЕ  
ВЫСОКИХ ЭНЕРГИЙ

1999

Применение ATM-технологии в экспериментальной физике  
высоких энергий

Рассмотрены вопросы применения ATM-технологий в экспериментальной физике высоких энергий. Приводятся примеры использования ATM-сетей в триггерных системах установки ATLAS. Описаны системы, в которых для передачи данных от детекторов к потребителям вместо общей магистрали применяются ATM-сети.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 1999

Перевод авторов

Kotov V.M., Nikityuk N.M.

E11-99-252

Use ATM-Technology in High Energy Physics Experiments

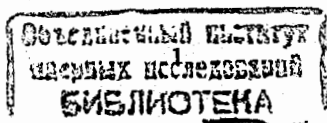
The questions of the use ATM-technology in high energy physics experiment are considered. The examples of the application of the ATM-nets in the trigger system of the setup ATLAS are considered. The systems where ATM-nets for the transmission data used instead a dataway are used described.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

## 1. Введение

Если рассматривать тенденцию использования средств вычислительной техники в экспериментах по физике высоких энергий, то можно отметить следующее. В 70–80-х годах широко использовались малые ЭВМ. В 80–90-х годах вследствие применения микропроцессоров получили распространение распределенные системы сбора и предварительной обработки физических данных (распределенные интеллектуальные системы), базируемые на тех или иных стандартах магистральных систем, где обмен данными происходит с помощью общей магистрали. И, наконец, начиная с середины 90-х годов наступает эра широкого применения коммуникационных вычислительных сетей [1]. Суть проблемы создания современных систем сбора и обработки данных заключается в том, что количество каналов регистрации в будущих экспериментах может составить  $10^8$  в одной установке. Учитывая высокую частоту встречи пучков на современных ускорителях встречных пучков и тем более на большом адронном коллайдере (LHC), магистральный способ передачи данных в системах сбора и обработки физических данных становится узким местом в триггерных системах, в которых требуется высокая пропускная способность (сотни Мбайт/с и более).

С целью выбора наиболее эффективной сетевой технологии в 1993 г. в ЦЕРН был предложен проект Nebulas (RD-31) [2], для исследования возможности использования в экспериментах физики высоких энергий ATM-технологии [3, 4]. Проект был направлен на изучение возможностей создания параллельной системы регистрации и обработки данных, получаемых на установке ATLAS. Основные направления исследований программы RD31: – детальное изучение возможностей использования выпускаемых промышленностью



коммутаторов в качестве строителей событий; – создание модели строителя событий; – разработка специализированных АТМ-интерфейсов.

Наряду с строителями событий АТМ-технология в экспериментальной физике высоких энергий используется для создания локальных и глобальных вычислительных сетей внутри систем сбора и обработки данных, вытесняя магистральные системы. АТМ-сети находят широкое применение для передачи в реальном времени не только данных с детекторов, но и мультимедийных приложений. В последующих разделах некоторые аспекты применения АТМ-технологии в экспериментальной физике будут рассмотрены более подробно.

## 2. Триггерная система установки ATLAS

На рис. 1 приведена структурная схема трехуровневой системы регистрации и отбора событий [5]. Частота поступления событий, которые будут регистрироваться с помощью детекторов, может составлять от 100 Мбайт/с до 10 Гбайт/с. По данным [6] при цикле 25 нс на внутренних трековых детекторах будет регистрироваться до 350 треков заряженных частиц и до 700 частиц может быть зарегистрировано в калориметрах. С помощью трехуровневой системы объем данных, поступающих от субдетекторов, может быть сокращен с сотен Гбайт/с до десятков Мбайт/с для последующей записи в системе массовой памяти. На рис.2 приведена стратегия отбора событий на каждом уровне триггерной системы [7]. В среднем столбце рисунка показано содержание физических процессов, данные которых обрабатываются на соответствующем уровне. Сигналы от субдетекторов с частотой 40 МГц с помощью оптоволоконных линий

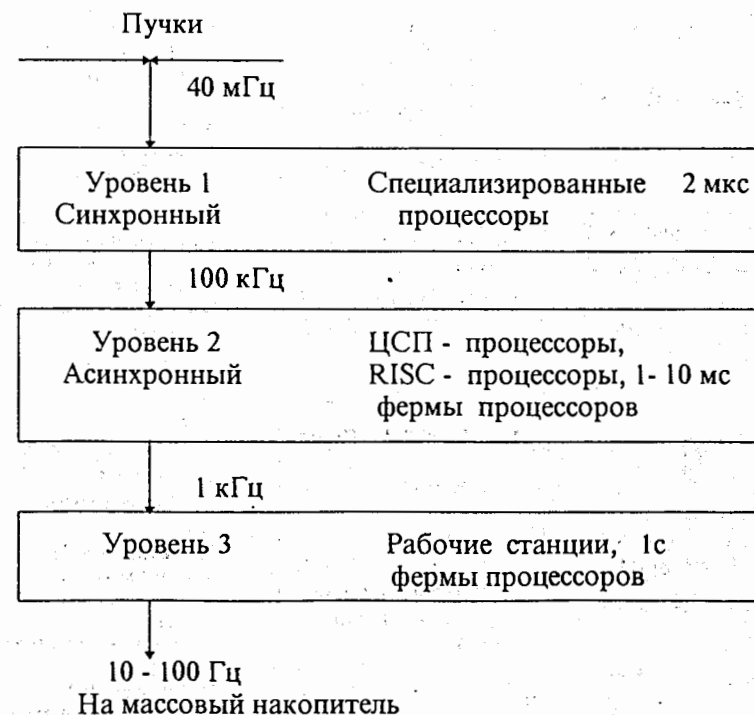


Рис.1. Блок-схема трехуровневой системы регистрации и фильтрации событий установки ATLAS

передачи поступают на первый уровень триггерной системы, который, по существу, представляет собой аналоговый конвейер. В каждом канале конвейера содержатся матрицы компараторов и аналоговые сдвиговые регистры для буферизации данных, благодаря чему система не имеет мертвого времени. На данном уровне используются также цифровые логические схемы и процессоры с табличными методами решения. Причем на первом уровне определяются области субдетекторов, в которых имеются указания на возможность регистрации полезного события (RoI-Region Of Interest). ROI-данные составляют 2% от максимально возможного количества регистрируемой информации в детекторе. При этом отбираются

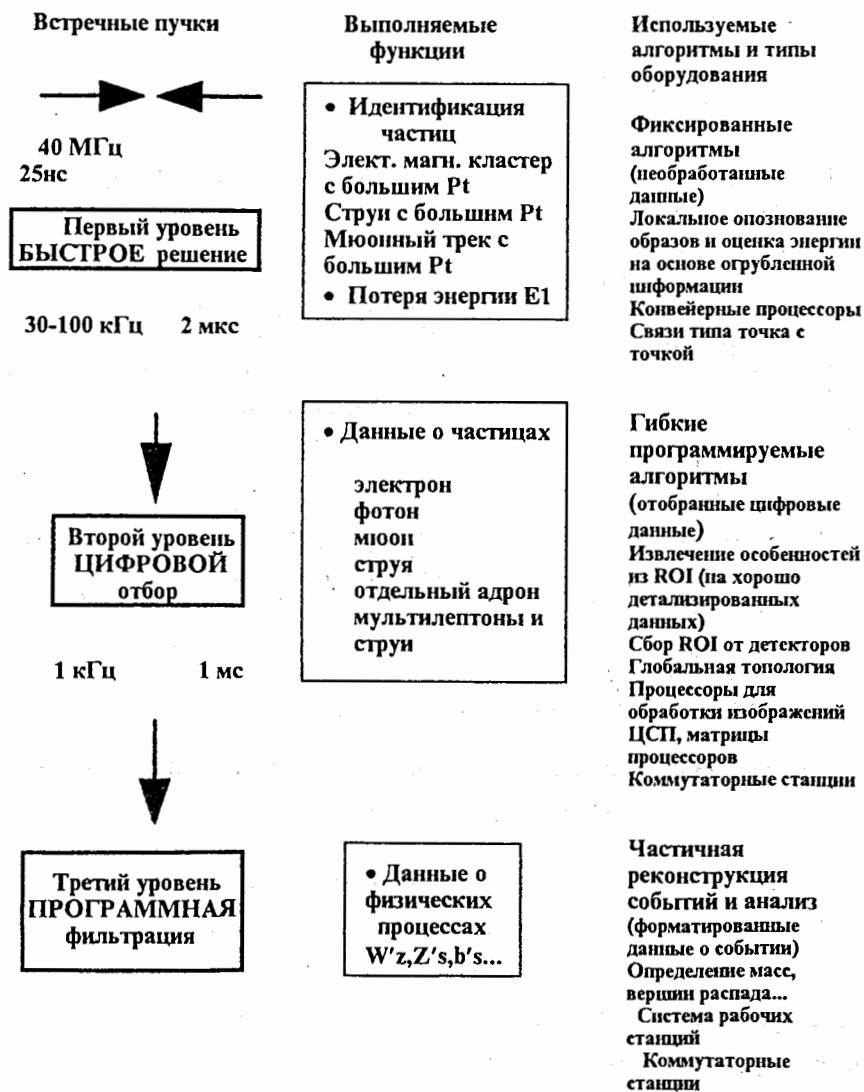


Рис.2. Стратегия отбора событий в эксперименте ATLAS. RoI - область интересов; Элект. магн. кластер - электромагнитный кластер

события, удовлетворяющие некоторым простым геометрическим критериям по топологии события и величинам энергий регистрируемых частиц. После оцифровки и мультиплексирования данные от принятых событий на первом уровне поступают на второй уровень триггерной системы с задержкой 2 мкс ( время решения триггера первого уровня ) и частотой 100 кГц. Причем ожидаемая полоса пропускания с учетом того, что на одно событие приходится свыше одного мегабайта данных, составляет порядка 100 Гбайт/с. Вот почему традиционные методы коммутации данных на основе магистральных систем не могут быть использованы в экспериментах на большом адронном коллайдере. На втором уровне величина задержки, определяемая временем решения по отбору полезных событий, составляет 1-10 мс, поскольку информация фильтруется по более сложным критериям. В качестве аппаратных средств используются в основном цифровые сигнальные процессоры промышленного изготовления.

С помощью триггерной системы третьего уровня, которая состоит из фермы мощных вычислителей, проводится анализ всего события. Затем данные о событии записываются на массовый накопитель. Частота записи данных на массовый накопитель составляет 10 - 100 Гц.

Как показывают исследования [8], обработка определенного события будет выполняться с помощью отдельного вычислительного узла. Причем для каждого события данные передаются к выбранному вычислительному узлу от нескольких сотен источников, определяемых внутри RoI. Поэтому уже на втором уровне триггерной системы используются матричные процессоры (процессорные фермы), а данные (RoI) от источников на фермы процессоров передаются с помощью построителя событий, который, по существу, представляет собой широкополосную сеть, содержащую сотни входов и выходов.

**Потоки данных.** На рис.3 приведена схема, иллюстрирующая потоки данных в триггерных системах второго и третьего уровней [12].

Сигналы, поступающие от субдетекторов, запоминаются в буферных ЗУ конвейерного типа (аналоговые или цифровые) на время работы триггерной системы первого уровня. Данные о событиях, принятых на первом уровне, с Read Output Driver (ROD) по оптическим линиям передаются в экспериментальный павильон и запоминаются в буферных запоминающих устройствах (ROB), количество которых составляет порядка 2000. Причем имеются отдельные буферы для калориметра, мюонного детектора и трековых детекторов. Данные, поступающие на

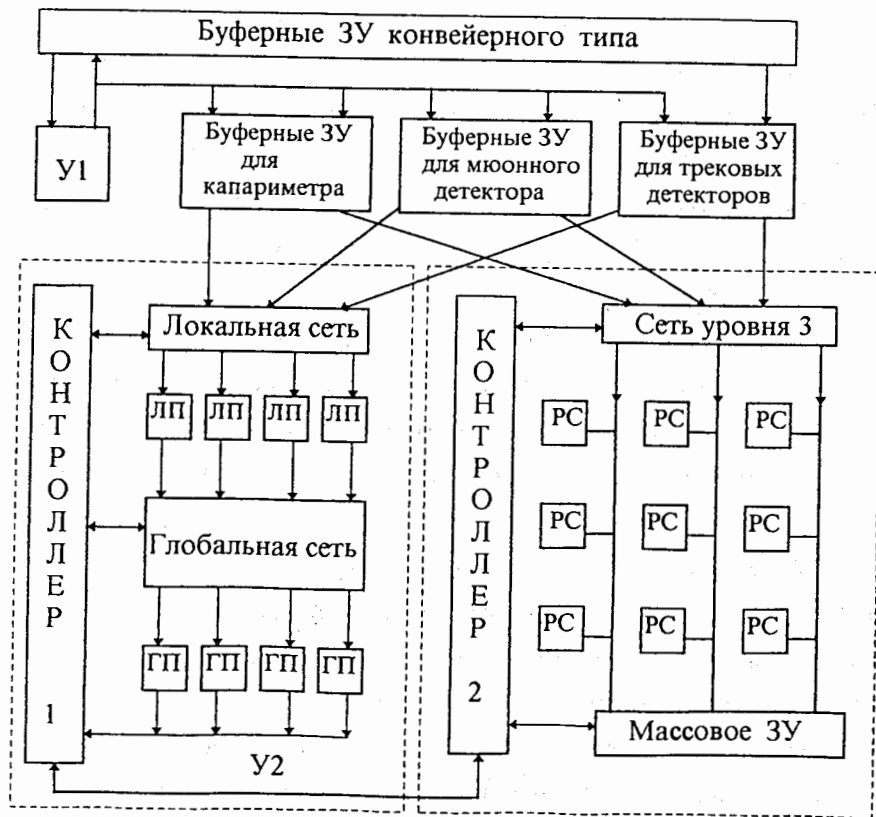


Рис.3. Блок-схема, иллюстрирующая потоки данных. У1, У2, У3 - триггерные уровни. ЛП - локальные процессоры, ГП - глобальные процессоры; РС - рабочие станции; ЗУ - запоминающее устройство

триггерную систему второго уровня, соответствуют RoI. Оценки показывают, что каждое событие содержит данные в среднем от 5 RoI. Причем данные из RoI для каждого субдетектора обрабатываются отдельно и параллельно с помощью локальных (ЛП) и глобальных (ГП) процессоров триггерной системы.

Следует отметить, что для каждого субдетектора данные от каждого RoI, как правило, поступают от разных ROD. Поэтому для объединения и передачи этих данных в процессор для дальнейшей обработки служит коммуникационная матрица, составляющая основу построителя событий. Результаты обработки, полученные с помощью локальных процессоров, передаются через глобальную сеть для дальнейшей обработки в глобальные процессоры в случае трехуровневой системы или остаются в одном и том же процессоре (что предпочтительно) в случае двухуровневой системы триггера, где второй и третий уровни обработки объединены в одной вычислительной ферме.

Таблица 1. Оценки требуемой полосы пропускания

СД	Объем события, Кбайт	ПП У2, Гбайт/с	ППУ3, Гбайт/с
Трековый детектор	770	10	6
Калориметр	400	5,6	3,2
Мюонный детектор	200	2,4	1,6
Всего	1370	18	10,8

Примечание: СД - субдетектор, ПП У2 -полоса пропускания уровня 2, ПП У3 - полоса пропускания уровня 3.

В таблице 1 приведены данные, характеризующие объем информации, регистрируемой в субдетекторах на одно событие [5].

Используя эти данные и учитывая, что частота событий, поступающая на входы триггерной системы второго и третьего уровней, равна 100 кГц и 1 кГц соответственно, можно оценить, что суммарная полоса пропускания триггерной системы должна составлять несколько десятков Гбайт/с.

### 3. Построители событий [5, 6–11]

**Блок-схема.** Как показали исследования, АТМ-сеть является наиболее удачной технологией для создания построителей событий. В частности, она обеспечивает необходимую полосу пропускания, может легко расширяться и работает надежно. Кроме того, стоимость оборудования неуклонно уменьшается.

По определению [6–8] построитель событий (event building), создаваемый на основе АТМ-технологии, представляет собой устройство, содержащее  $N$  источников и  $M$  приемников данных, соединенных между собой сетью. Основу такой сети составляет АТМ-коммутирующая станция. Другими словами, построитель событий включает в себя коммутационную сеть с соответствующими устройствами управления, определенное количество источников данных, заданное количество приемных модулей, в которых запоминаются фрагменты событий. В качестве приемников данных обычно используются буферные ЗУ. Для каждого события приемники собирают данные от субдетекторов и восстанавливают полную топологию событий с помощью мощных процессоров. Коммутационная сеть одновременно поддерживает параллельные потоки данных, так что множество событий восстанавливается одновременно в конкурентном режиме.

На рис.4 приведена блок-схема типичного построителя событий. Данные на входы источников  $S_1 - S_N$  поступают от триггерной системы

первого уровня. Виртуальные соединения, которые обозначены штрихованными линиями, открыты постоянно. После появления триггерного сигнала на отобранное событие контроллер инициирует работу всех источников в радиовещательном режиме. При этом каждый источник с помощью коммутатора посылает фрагменты события через виртуальные соединения на обозначенные приемники  $P_1 - P_M$ . Для передачи данных используется протокол ААL5. Причем порядок поступления ячеек, содержащихся в ААL5-пакете и поступающих на входы, такой же, как и порядок ячеек, формируемый на передающей стороне. Однако порядок поступления фрагментов на приемники, относящихся к различным событиям, заведомо не определен и зависит от наличия свободного виртуального пути или очередности, устанавливаемой в источниках. В конце концов все фрагменты данного события окажутся в одном заданном приемнике данных.

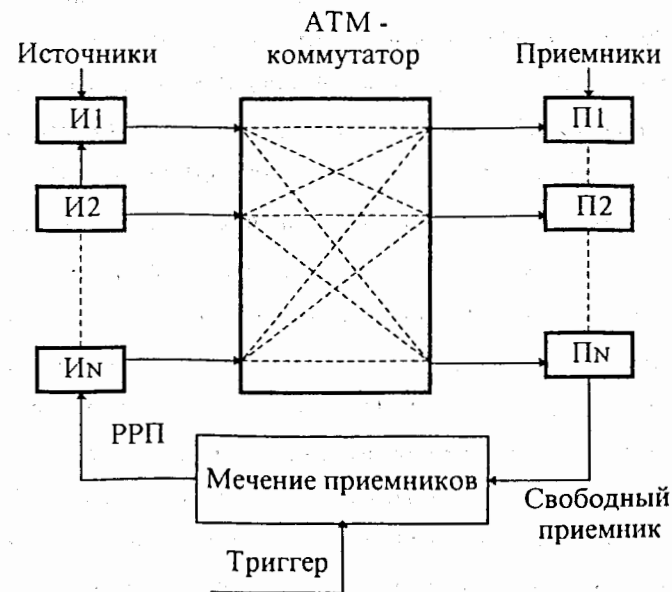


Рис.4. Блок-схема типичного построителя событий. И<sub>1</sub>-И<sub>N</sub> - источники; П<sub>1</sub> - П<sub>N</sub> - приемники

**Основные параметры и проблемы.** Построители событий характеризуются рядом параметров.

1. Задержка построителя событий равна времени между поступлением фрагмента события на вход построителя событий и формированием его на приемнике.

2. Загрузочный фактор представляет собой максимальную часть агрегатной полосы пропускания коммутатора, которая используется для построения события.

3. Частота работы построителя определяет время, в течение которого происходит формирование события или его фрагмента.

4. Максимально возможная очередь ячеек на входах и выходах коммутаторной станции. Эта характеристика определяет емкость необходимой памяти на входах коммутатора.

5. Количество входных/выходных портов.

Следует отметить, что коммутаторные станции в основном выпускаются для телекоммуникационных систем и поэтому в них не учитываются особенности передачи данных в экспериментах по физике высоких энергий. Суть проблемы заключается в том, что для каждого события все источники данных могут посылать в одно и то же время фрагменты событий на один и тот же приемник. В результате возникают заторы даже в случае большой суммарной пропускной способности коммутаторной станции. Для устранения заторов и состязательности ячеек при коммутации с помощью ATM-коммутаторов RD-31-коллорабация предлагает использовать два способа: выравнивание частоты поступления ячеек на входах коммутирующей станции (traffic shaping) и контроль потока ячеек внутри станции (internal flow control) [2, 5, 7]. Первый способ предполагает ограничение средней пропускной способности таким образом, чтобы суммарная пропускная способность на входах приемников не превышала номинальной пропускной способности выходных портов, и исключение, по возможности,

кластеров так, чтобы данные на приемники поступали равномерно. Другими словами, необходимо, чтобы в каждом соединении номинальная пропускная способность была равна  $1/N$ . При этом выравнивание частоты поступления ячеек заключается в управлении потоками данных от источников: во-первых, в контроле скорости передачи ячеек по отдельным виртуальным соединениям; во-вторых, в измерении временной зависимости величины загрузки и изменения типа коммутации. При использовании второго способа (internal flow control) в случае возникновения затора коммутаторная станция посылает сигнал источнику, в соответствии с которым он должен уменьшить частоту поступления ячеек. С целью повышения эффективности работы построителя событий могут использоваться оба способа одновременно.

#### Модель построителя событий на базе ATM-технологии [5, 7-12]

На рис.5 приведена блок-схема модели, созданной на основе восьмипортового ATM-коммутатора с пропускной способностью 155Мбит/с. Прототип содержит ATM-интерфейс (см. ниже), выполненный в стандарте VME на основе RISC-процессора. Он

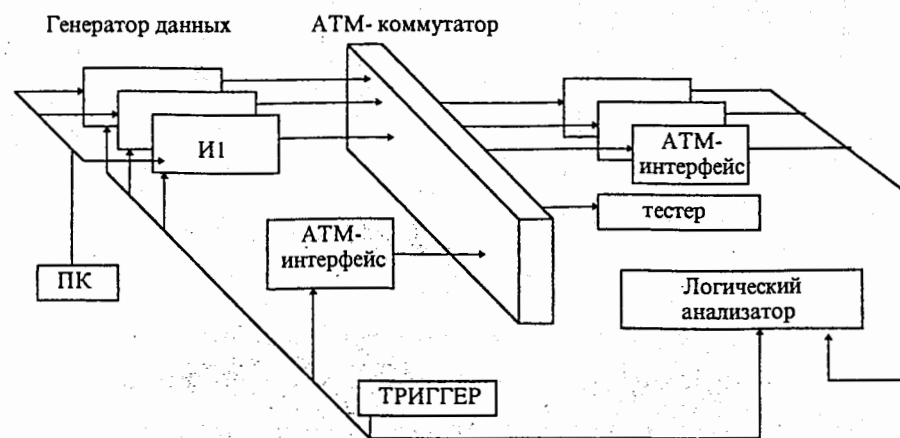


Рис.5. Блок-схема модели построителя событий



используется качестве источников и приемников данных. Интерфейс представляет собой обычную VME-плату, с помощью которой реализуются уровни ATM-протокола, AAL5, ATM и физический уровень. Для этих целей используются специализированные микросхемы. Причем с помощью физического уровня реализуется 155Мбит/с SDH-протокол. Для передачи данных используется одномодовое оптоволокно. Анализатор и тестор используются для контроля работы построителя событий и измерения внутренних времен задержки. Кроме того, в состав прототипа входит персональный компьютер, выполняющий функцию генератора данных.

**Протоколы построителя событий.** На рис.6 приведена структура

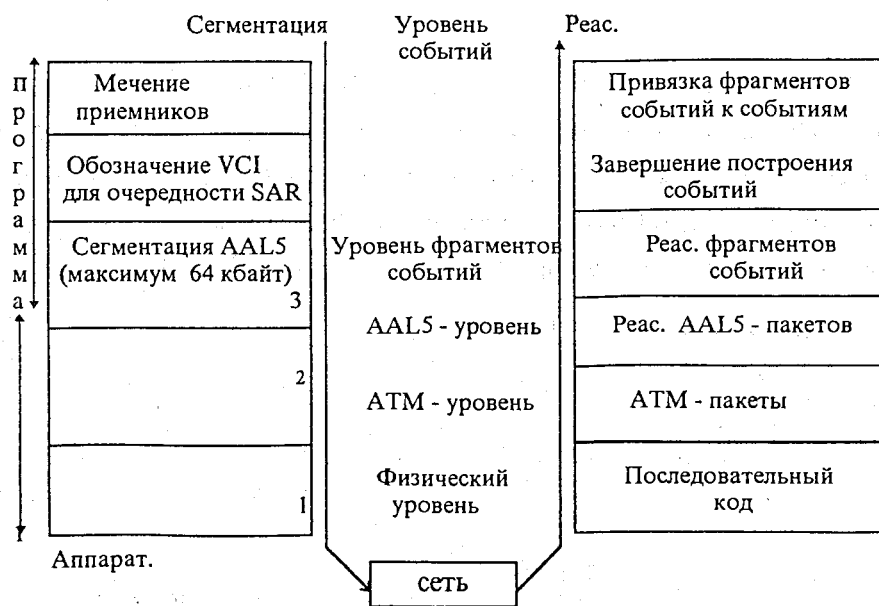


Рис.6. Структура протокола построителя событий. Реас. - реассемблирование; SAR - система ассемблирования и реассемблирования; Аппарат. – аппаратная часть

протокола построителя событий. Уровни 1 - 3 являются стандартными AAL5, ATM и физический уровень (SONET/SDH) соответственно. Протокол AAL5 используется для передачи фрагментов событий и служебной информации. Как отмечалось выше, AAL5-стандарт определяет передачу данных блоками емкостью 64 кбайт. Сегментация и реассемблирование ATM-ячеек выполняются с специализированной микросхемы. Уровень "События" выполняет функцию объединения вместе фрагментов данного события. На этом уровне происходит формирование нескольких событий одновременно в конкурентном режиме и определяется время окончания сбора данных для каждого события. В [12–13] модель построителя событий создана на базе ATM-коммутатора и быстрых рабочих станций, которые используются в качестве источников и приемников информации. Кроме того, в состав построителя событий входят также генератор событий, система запуска и контроля на основе сети Ethernet .

**Применение транспьютеров.** Весьма перспективным направлением в создании построителей событий является применение транспьютерных сетей. Одна из таких сетей, в которой используется высокопроизводительный транспьютер T9000 и ассоциированный с ним асинхронный коммутатор C104 [14], имеющий задержку 1 мкс и пропускную способность 264 Мбайт/с в обоих направлениях.

#### ATM-сети и распределенные системы сбора и обработки

**экспериментальных данных.** В будущих экспериментах количество каналов регистрации будет возрастать. Кроме того, сложилась тенденция в создании распределенных систем регистрации и обработки данных. Пропускная способность такой системы должна составлять сотни и более Мбайт/с. Разработан ряд установок и распределенных систем, в которых для передачи данных на большие расстояния в

реальном времени используются сети. Пожалуй, одна из первых установок, в которой для обмена данными между отдельными подсистемами вместо общей магистрали используется локальная сеть, описана в [15]. Часть аппаратуры с детекторами расположена на стратостате. Данные на землю передаются с помощью радиосигналов. Информация форматируется в виде пакетов так, как это рекомендуется в стандарте X.25. Наземная аппаратура состоит в основном из персональных компьютеров и рабочих станций, объединенных с помощью локальной сети. В [16] описана более совершенная установка, в которой для реализации протокола используется специализированная микросхема. На базе этой микросхемы построен контроллер, который находится в каждой подсистеме. Скорость передачи данных составляет 4 Мбит/с. Использование локальной сети позволяет каждой подсистеме, входящей в наземную часть установки, работать независимо одна от другой или отключаться от сети в случае необходимости. Кроме того, система легко расширяется.

ATM-сети становятся неотъемлемой частью создаваемых экспериментальных установок. Как показано в [17, 18], объем данных при регистрации и скорость их обработки (DAQ - Data Acquisition System) можно увеличить, по крайней мере на порядок, путем использования ATM-технологии. Кроме того, поскольку система получается распределенной, повышается эффективность сотрудничества больших коллабораций: появляется возможность использовать общую информацию без необходимости копирования данных; данные поступают в реальном времени, практически независимо от расстояния до ускорителя и от конечных пользователей; появляется возможность контролировать работу установки в режиме удаленного доступа; проводить видеоконференции и текущее обсуждение получаемых результатов; в будущем появится возможность

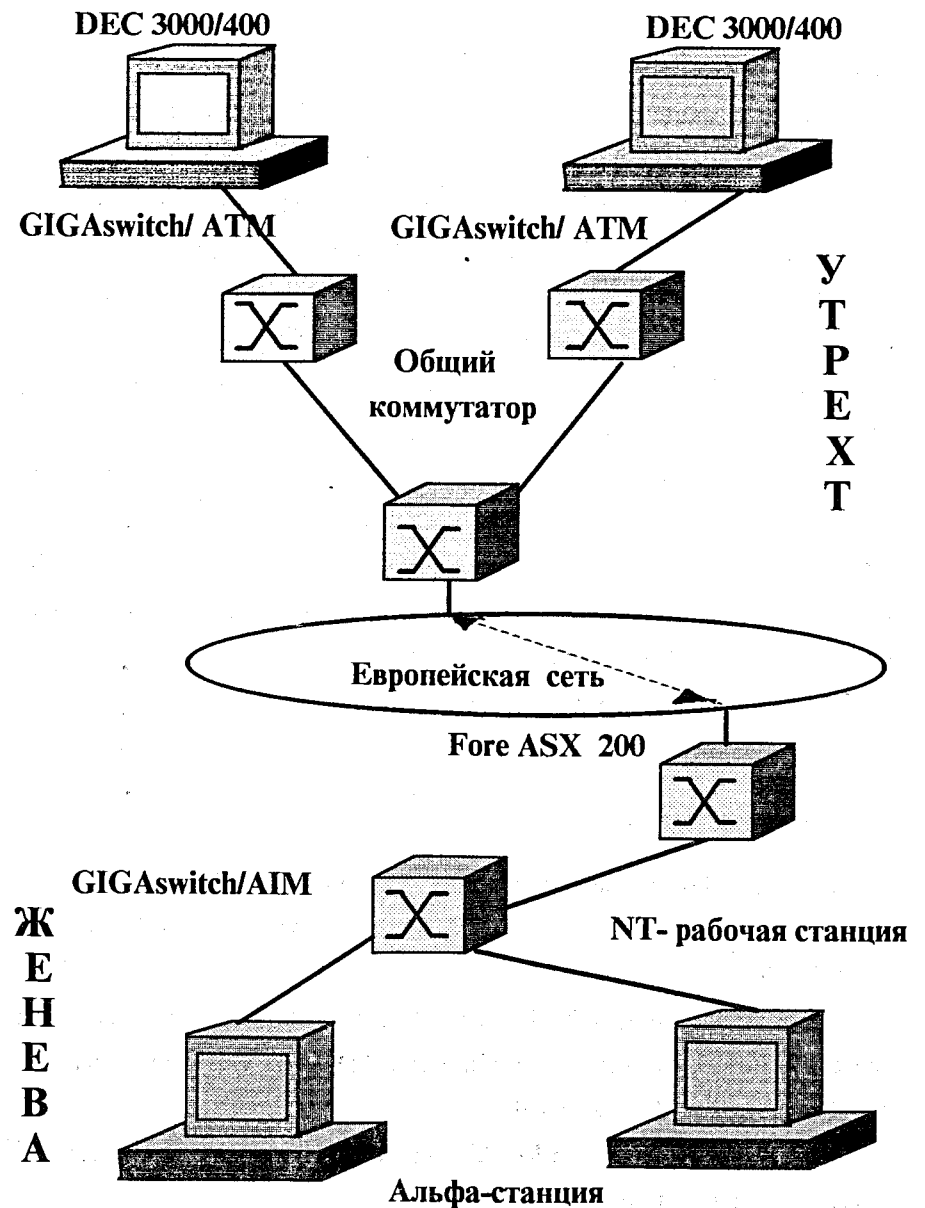


Рис.7. Экспериментальная блок-схема сети ATM Утрехт - Женева.

DEC 3000/400 - рабочие станции фирмы "DEC"; Gigaswitch/ATM - быстрый ATM-коммутатор; Fore ASX 200 - ATM-коммутатор фирмы "Fore"

проведения экспериментов без присутствия физиков непосредственно на экспериментальной установке.

В рамках национальной сети SURFnet (Голландия) реализуется проект "Удаленная система регистрации, обработки и анализа физической информации." В состав распределенной сети сбора данных по состоянию на 1996 г. входят лаборатория Ван Граафа (Утрехт), детектор GANIL (Франция) и нейтронный калориметр (Амстердам). Следующим этапом развития проекта будет включение в общую сеть установки для исследования плазмы и обеспечение связи с ЦЕРНом [19] (рис.7).

Таким образом, использование ATM-технологии и мощных рабочих станций является весьма перспективным в экспериментальной физике. Однако в таких сетях необходимы многоуровневые системы управления в связи с необходимостью передачи разнородных мультимедийных потоков данных и для исключения заторов, которые резко снижают эффективность системы [20].

#### Список литературы

1. Bu P. Le Du. Requirements for Future Experiments. In: Proc. of the Workshop " Applied of ATM in High Energy Physics". Paris 21-22 sept. 1998. P.13-23.
2. Christiansen J., Dufey J.P., Letheren M. et al. The "Nebulas" (RD-31) project. In: Proc. of the Int. Conference RTD'94. JINR, Dubna, Russia, June 27-July 1, 1994. P. 299-310.
3. Котов В.М., Никитюк Н.М. ATM-технология. Сообщение ОИЯИ. РМ-99-13. Дубна 1999.
4. Назаров А.Н., Симонов М.В. ATM: Технология высокоскоростных сетей. Из-во ЭКО-ТРЕНДЗ. Москва, 1998, 234 с.

5. Calvet D., Djidi K., Du Le et al. A Study of Performance Issues of the ATLAS Event Selection System Based on ATM Switching Network. IEEE Trans. on Nuclear Sci. 1996. V. 43. No. 1. P. 90-98.
6. Mapelly L. Architecture of Future Data Acquisition Systems. 1992 CERN School of Computing. CERN 93-03. Geneva, 1993. P. 237-335.
7. Costa M., Dufey J.P., Letheren M. et al. ATM-Based Event Building. ATLAS Internal Note. DAQ-No-24. Geneva, 1994.
8. Letheren M., Christiansen J., Mandjavidze I. et al. An asynchronous Data-Driven Event-Building Scheme Based on ATM Switching Fabrics. IEEE Trans. on Nucl. Sci. 1994. V. 41. No. 1. P. 257-266.
9. Costa M., Dufey J.P., Letheren M. et al. An ATM-Based Event Builder System. In: Proc. First Workshop on Electronics for LHC Experiments. Lisbon, sept. 11-15, 1995. P. 340-344.
10. Costa M., Dufey J.P., Letheren M. et al. Lessons from ATM-Based Event Builder Demonstrators and Challenges for LHC-Scale System. In: Proc. Second Workshop on Electronics for LHC Experiments. Balatonfured, Hungary, sept. 23-27, 1996. P. 208-214.
11. Costa M., Dufey J.P., Letheren M. et al. Results from an ATM-Based Event Builder Demonstrator. IEEE Trans. on Nucl. Sci. 1996. V.43. No.3. P. 1814-1820.
12. Calvet D., Hubbard R., Huet M. et al. Operation and performance of ATM-Based Demonstrator for Sequential Option of the ATLAS Trigger. IEEE Trans. on Nucl. Sci. 1998. V.45. No.4. P. 1793-1797.
13. Calvet D. Evaluations of ATM for the ATLAS Trigger and Data Acquisition System. In Proc. of the Workshop "Application of ATM in High Energy Physics." Paris 21-22 sept. 1998. P. 2/16-16/16.
14. Dobinson R.W., Francis D., Heeley R. et. al. Triggering and Event Building Results using the C104 Paket Routing Chip. Nucl. Instr. and Methods. 1996. V. A376. No. 1. P. 59-66.
15. Imori M., Yoshimura K., Ueda I. et al. A Remote Console System for

- Ballon-Borne Experiment. IEEE Trans. on Nucl. Sci. 1992. V. NS-40. No. 4. P.913-915.
16. Imori M., Yoshimura K. Ueda I. et al. A Local Network Integrated into a Ballon-Borne Apparatus. Nucl. Instr. and Methods. 1998. V. A414, No.1. P.194-204.
17. C.Th.A.M. de Laat, Kuijer P.G., Othuis H.P. et al. A Distributed Data Acquisition System by ATM Networks a Pilot Study. IEEE Trans. on Nucl. Sci. 1996. V. 43. No. 1. P. 123-127.
18. C.Th.A.M. de Laat, Kuijer H.P., Othuis H.P. et al. Experiences with the Application of LAN Emulation in a Data Acquisition System. IEEE Trans. on Nucl. Sci. 1997. V. 44. No. 4. P. 1635-1638.
19. Andree H.M.A., Habets J., Koopmans M. et al. Virtual Control Room, the "REMOT" Project, Networking Pilot Studies. IEEE Trans. on Nucl. Sci. 1998. V. 45. No. 4. P.1999-2003.
20. Nomachi M., Sygaya Y., Togawa H. et al. Performance Measurement of Mixed Data Acquisition and LAN Traffic on a Credit-Based Flow-Controlled ATM Network. IEEE Trans. on Nucl. Sci. 1998. V. 45. No. 4. P. 1854-1858.

Рукопись поступила в издательский отдел  
21 сентября 1999 года.