Интерфейсы и периферийные устройства

Лекция 4-1. Шина PCI Express

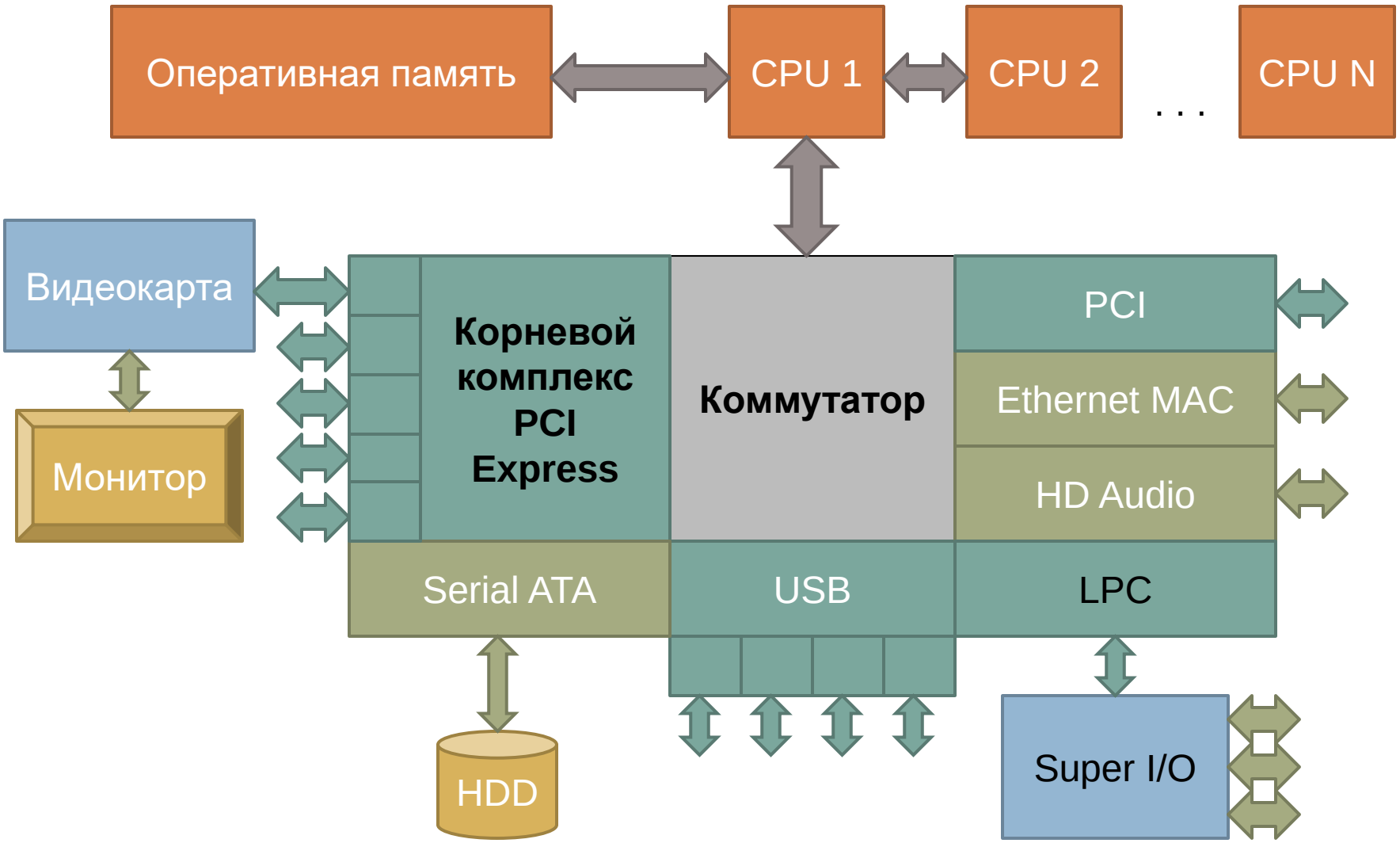
Предпосылки создания

Общая характеристика

Коммутационная фабрика

Уровни протокола PCI Express

Архитектура современного ПК



Концепции PCIe

масштабируемость и **универсальность**.

- *Универсальность* шины PCI Express должна заключаться в том, чтобы она заменила шину, связывающую северный мост чипсета с графическим адаптером, шину, объединяющую северный и южный мосты чипсета, а также PCI-шину.
- *Масштабируемость* шины PCI Express состоит в том, что шина позволяет наращивать пропускную способность от 2,5 Гбит/с вплоть до 10 Гбайт/с (80 Гбит/с). Для сравнения отметим, что пропускная способность шины PCI-X с частотой 133 МГц составляет 1,06 Гбит/с.

Почему PCIe?

- Суть проблемы заключается в том, что со временем появляется все больше устройств ввода-вывода, требования по быстродействию которых не соответствуют возможностям шины PCI.
- Скорость & частота
- Коммутация каналов & пакетов
- Еще один недостаток шины PCI состоит в чрезмерных габаритах плат.

Шина PCI Express

Шина PCI Express (проект Arapahoe) была разработана в 2002 году как универсальный периферийный интерфейс системного уровня.

Первая общепринятая спецификация имеет версию 1.0a, она была принята комитетом PCI SIG в 2003 году.

Позднее была принята спецификация 1.1,

В 2007 году одобрена спецификация 2.0.

Версия 3.0 - в 2010 году.

При разработке PCI Express особое внимание было уделено **совместимости с PCI на уровне механизма конфигурирования, программного доступа и поддержки со стороны ОС и драйверов.**

При этом требовалось **сохранить или уменьшить стоимость реализации при значительном улучшении всех характеристик, прежде всего пропускной способности.**

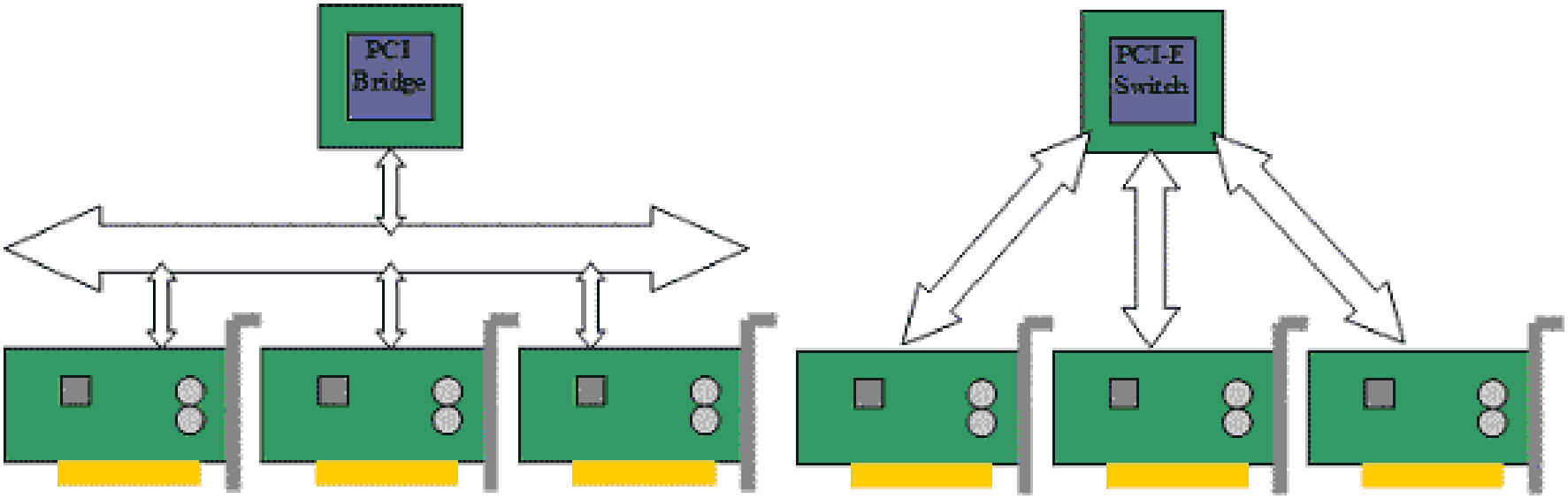
PCI Express Link

Вместо шинного соединения PCI в PCI Express применена схема **объединенных через коммутаторы двухточечных каналов связи** между устройствами и портами.

Соединение (Link) – это пара встречных симплексных каналов - передающий (Transmitting) и принимающий (Receiving). Каждый канал является низковольтной дифференциальной парой сигналов (LVDS, <=2,6 v). Сигналы идут одновременно в противоположных направлениях.

Скорость соединения (Signaling Rate) устанавливается в начале работы шины; определены две скорости – 2.5 Гбит/с и 5.0 Гбит/с (PCIe 2.0).

Сравнение топологий PCI и PCI Express



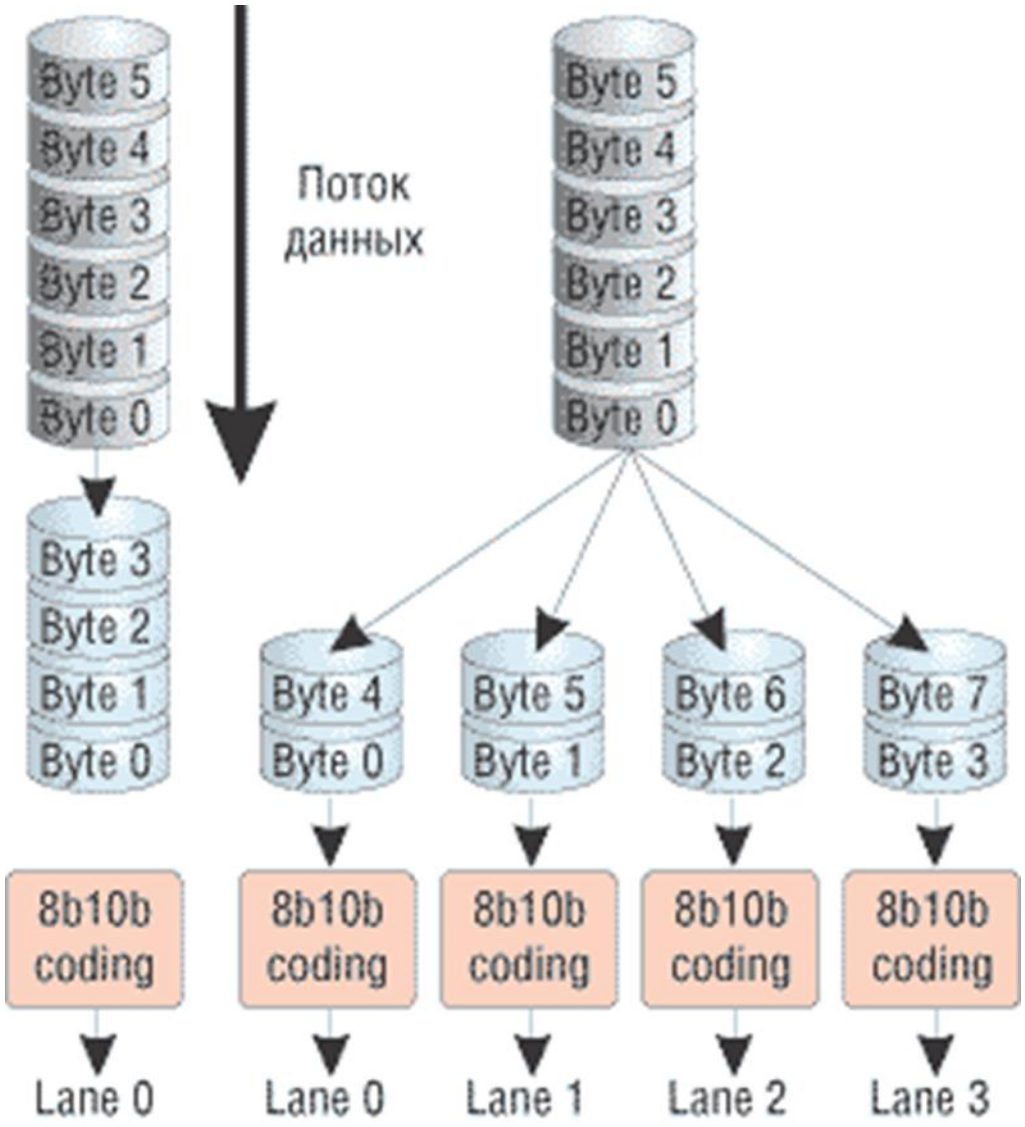
Масштабирование соединения

Соединение может агрегировать несколько линий.

Спецификация предусматривает следующие конфигурации соединения: x1, x2, x4, x8, x12, x16, x32.

Количество дифференциальных пар на прием и передачу должно быть одинаково, **несимметричные соединения невозможны.**

Данные по разным линиям передаются побайтно, общий поток делится на блоки, кратные количеству линий



Пропускная способность

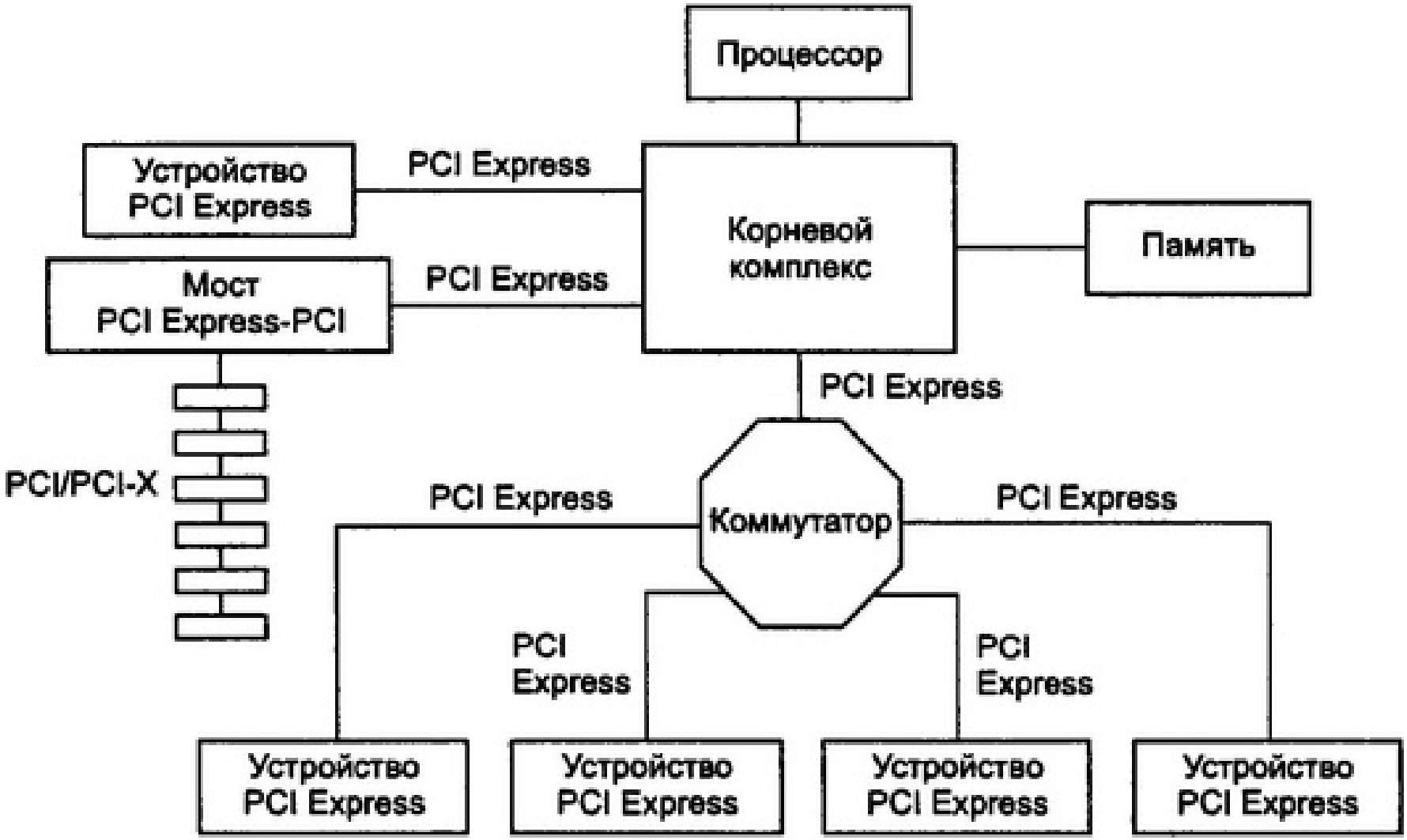
Для расчета пропускной способности соединения необходимо учесть то, что в каждом соединении передача дуплексная, а также учесть применение кодирования 8B/10B (8 бит в 10). Таким образом, полная (дуплексная) пропускная способность соединения x1 составляет 2,5 x 2 x 0,8 / 8 = 0,5 ГБ/с (эффективная скорость передачи информации в одном направлении - только на запись или только на чтение - соответственно в 2 раза ниже).

Вся последовательность передаваемых данных распределяется на имеющиеся линии, *передача параллельная, но не синхронная*. В примере: если имеется 4 линии, то 0-й байт блока данных передается по 1-й линии, 1-й - по 2-й, и т. д., а 4-й - снова по 1-й. Соответствующая пропускная способность возрастает строго пропорционально, так что для x4 максимальная скорость передачи уже 2 ГБ/сек.

В первую очередь PCI Express используется для подключения дискретных видеокарт. Звуковая карта SUS Xonar DX.

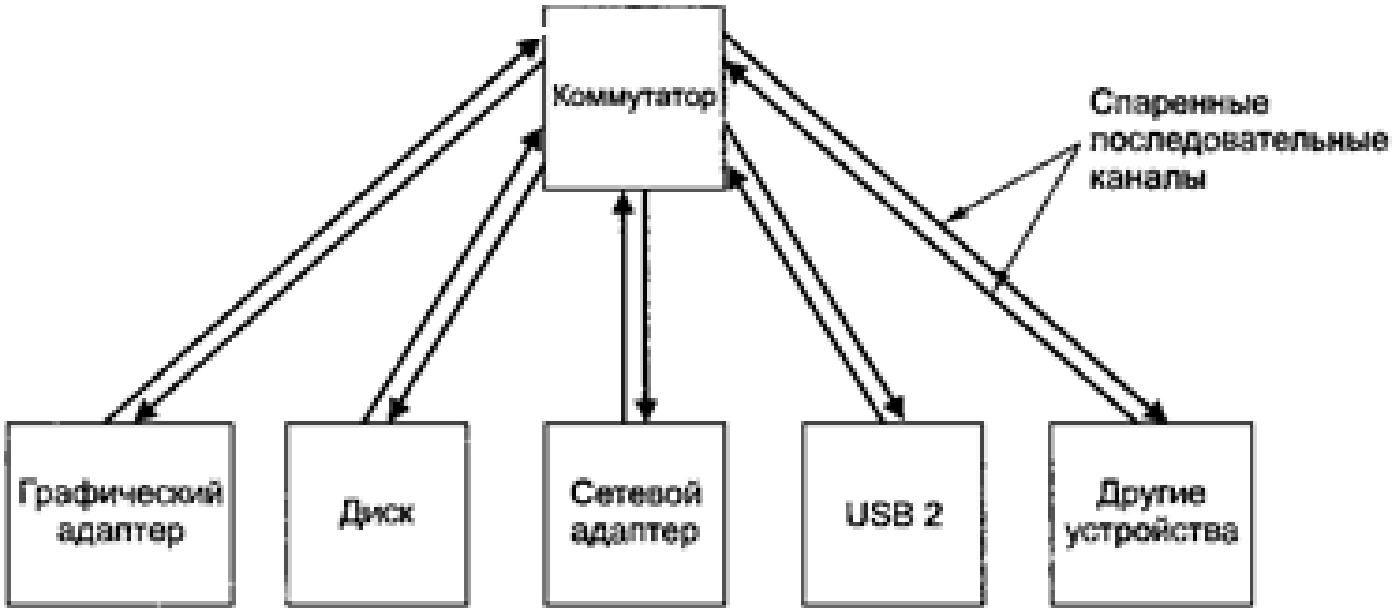
SSD накопитель OCZ Z-Drive R4 Enterprise

Коммутационная фабрика PCI Express



Использование

В первую очередь PCI Express используется для подключения дискретных видеокарт (графических адаптеров).



Другие устройства : звуковая карта (SUS Xonar DX);
SSD накопитель (OCZ Z-Drive R4 Enterprise)
…

Корневой комплекс (RC)

Это аналог главного моста (Host Bridge) в шине PCI – инициализирует и управляет фабрикой). Он отвечает за связь с процессором и системной памятью, а также за конфигурирование всей фабрики.

RC содержит несколько портов PCI Express (Root ports), которые могут (необязательно) взаимодействовать между собой посредством виртуального коммутатора. К каждому из портов RC может подключаться:

- коммутатор (switch),
- мост для другой шины (напр., PCI),
- конечное устройство (Endpoint).

RC отвечает за конфигурационные циклы, может выполнять циклы доступа к портам и пространству памяти.

RC может запрашивать блокированные (Locked) операции, но не может отвечать на запросы с блокировкой.

Конечное устройство (Endpoint)

Каждое конечное устройство подключается к порту либо RC, либо коммутатора.

Устройство выполняет транзакции от своего имени либо от имени подключенной к нему шины, устройства или контроллера другого интерфейса.

Устройства могут быть полноценными и устаревшего типа (Legacy).

Полноценное устройство:

- Не работает через порты – только через диапазон памяти
- Не работает с блокированными запросами
- Поддерживает 64-битное адресное пространство по умолчанию
- Поддерживает механизм прерываний MSI, причем с 64-битным пространством
- Имеет расширенное пространство конфигурирования

Механизм конфигурирования

Позаимствован у PCI-X 2.0. Стандартный способ доступа – через конфигурационный цикл – сохранен для совместимости. Полное конфигурационное пространство каждого устройства занимает 4 Кб.

Для упрощения доступа к конфиг. регистрам предусмотрен механизм их отображения на пространство памяти. По заданному базовому адресу находится пространство для всех возможных устройств в рамках системной шины

Memory Address[62]	PCI Express Configuration Space
A[27:20]	Bus Number
A[19:15]	Device Number
A[14:12]	Function Number
A[11:8]	Extended Register Number
A[7:2]	Register Number
A[1:0]	Along with size of the access, used to generate Byte Enables

Порт PCI Express

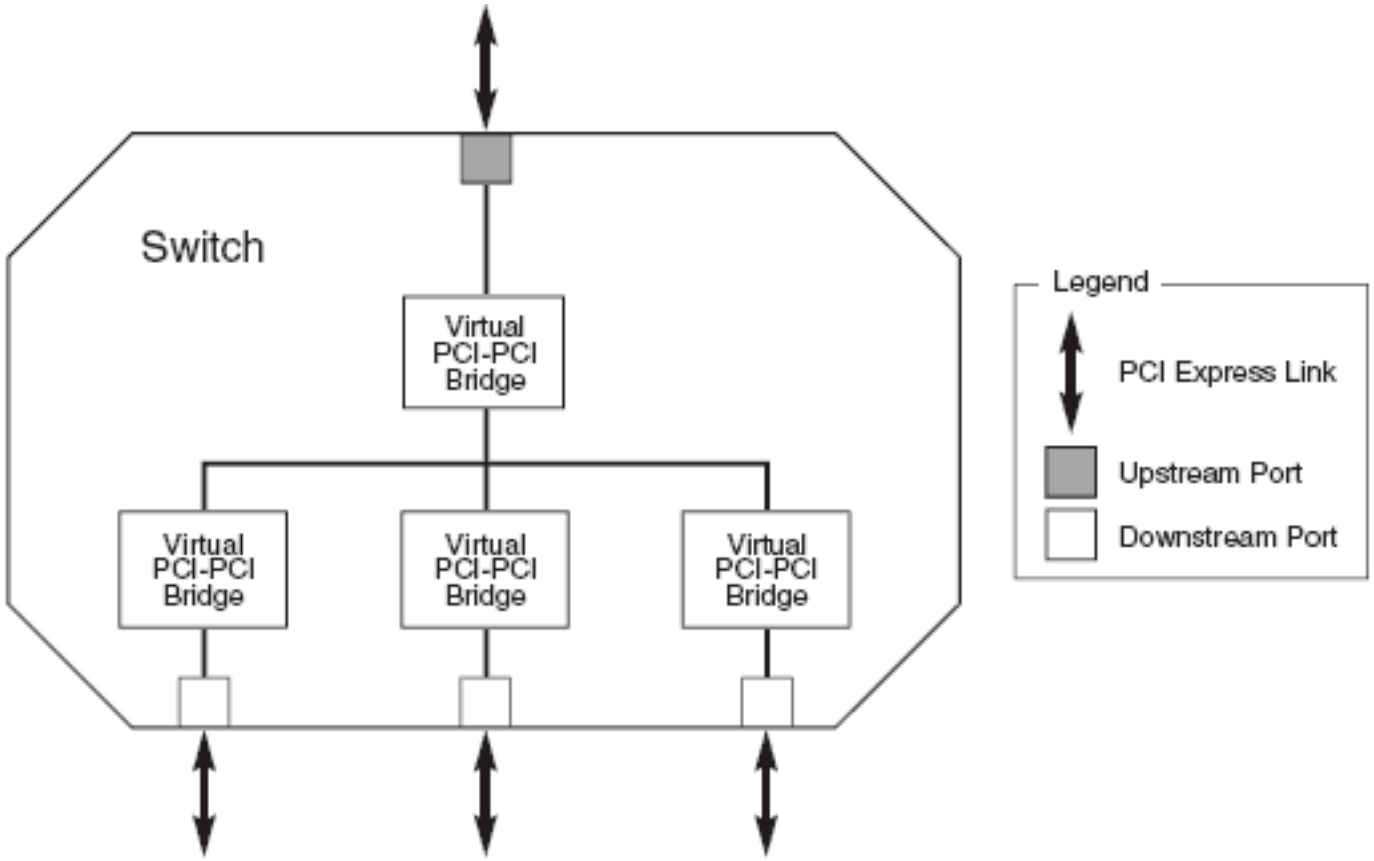
Порт – это логическая точка подключения соединения (Link), которая отвечает за управление линиями, сборку в пакеты исходящих данных и разборку входящих. Портами оснащен RC и коммутаторы (если они имеются).

С точки зрения программирования порт представляет собой виртуальный мост PCI-PCI, а его Link – виртуальную подчиненную (вторичную) шину PCI.

Все порты делятся на корневые (принадлежат RC), нисходящие и восходящие (последние – только у коммутаторов).

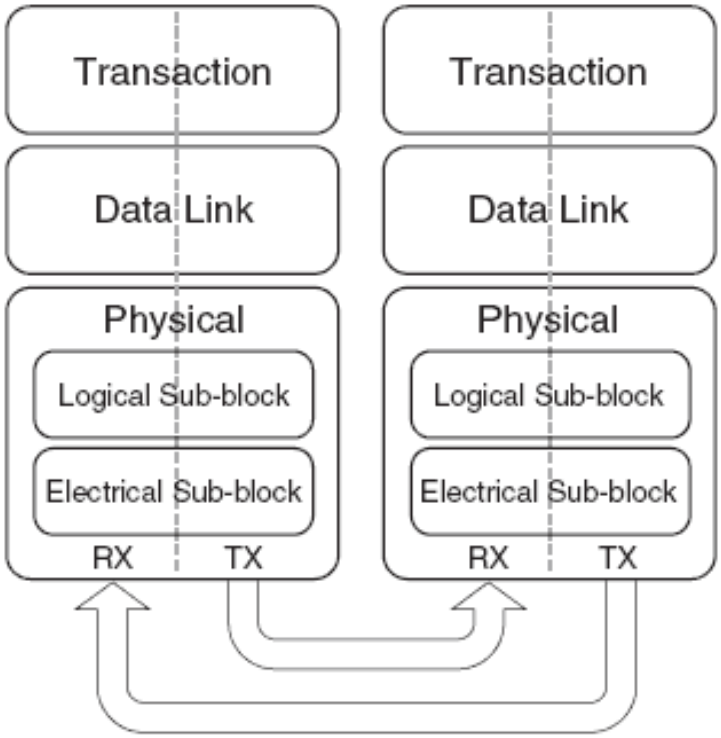
Коммутатор PCI Express

Коммутатор служит для расширения количества подключаемых устройств, это аналог моста дополнительных шин PCI. Программно коммутатор представляет собой набор мостов PCI-PCI. Один из портов коммутатора ведет к порту RC или другого коммутатора.

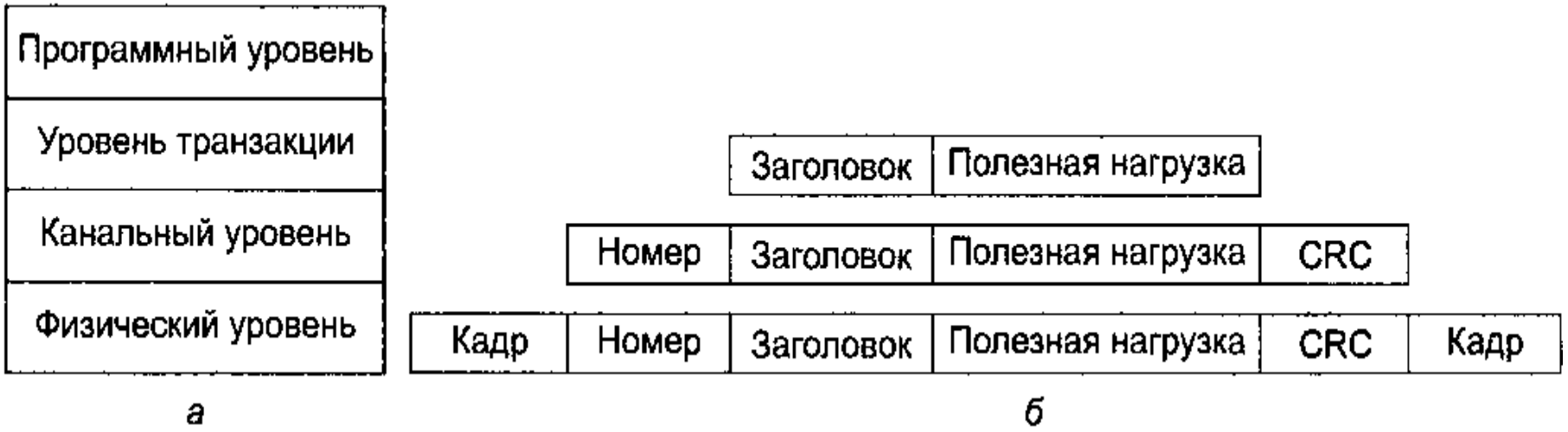


Уровни протокола PCI Express

В отличие от PCI протокол PCI Express условно разделен на уровни, без уточнения способов их реализации. Уровней всего три, на каждом выполняется сборка и разборка пакетов и их обрамление необходимыми заголовками и контрольными суммами. Не все пакеты относятся к уровню транзакций, существуют пакеты только канального уровня, служащие для управления.



Стек протоколов и форматы пакетов



а *б*

Уровень транзакций

Этот уровень отвечает в основном за выполнение операций чтения и записи в память либо в порты ввода-вывода.

Все транзакции, требующие ответа (обычно чтение), выполняются как расщепленные (Split): их инициатор получает статус запросчика (Requester), а целевое устройство – статус исполнителя (Completer).

Уровень транзакций отвечает и за управление потоком, реализованное на основе механизма кредитов.

На уровне транзакций поддерживается 4 адресных пространства:

- Памяти (основное) (при выполнении стандартных операций чтения и записи)
- Портов в-в (для совместимости) (для адресации регистров устройств)
- Конфигурационное (для инициализации системы и т. д.)
- Пространство сообщений (Message Space) (для отправки сигналов, прерываний и т. д. - для эмуляции сигналов шины PCI (INTx#, PME# и др.) – т.н. «виртуальные провода».

Адресные пространства

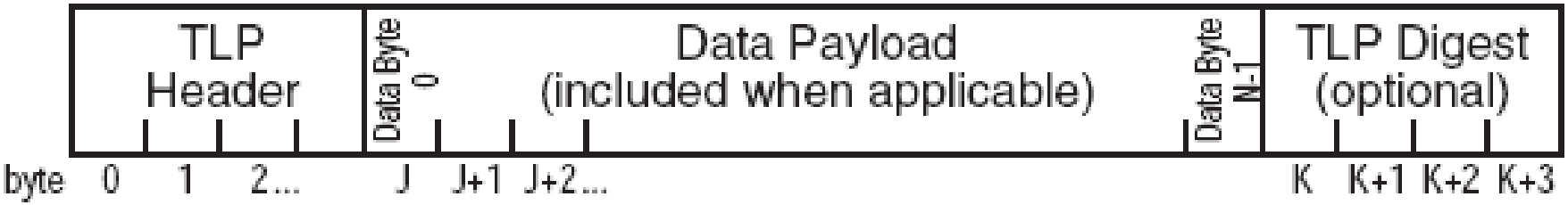
- Пространства памяти и ввода-вывода аналогичны традиционным - тем, что реализованы в современных системах.
- В конфигурационном пространстве возможна реализация разного рода механизмов, например автоматического конфигурирования (PnP).
- Пространство сообщений принимает на себя функции многочисленных ныне управляющих сигналов. Обойтись без этого пространства нельзя, ведь в PCI Express отсутствуют предусмотренные в шине PCI линии управления.

Пакеты уровня транзакций

Пакеты шины PCI Express оптимизированы для передачи по высокоскоростным последовательным линиям. Они имеют переменный формат, в том числе длину, чтобы исключить передачу незадействованных полей.

Первым передается наиболее значимый байт, обычно байт №0, чтобы приемное устройство могло начать его обработку до прихода остальных байтов.

Формат (обобщенный) пакета TLP следующий:



Длина пакета выровнена по границе dword.

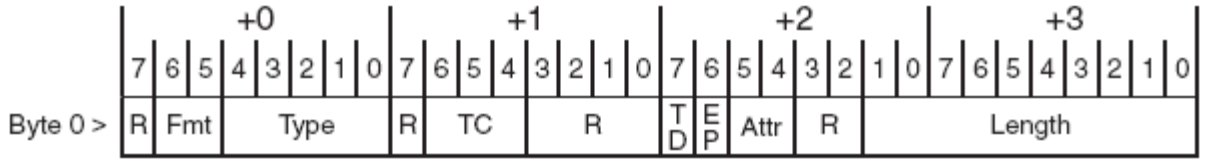
Код CRC обеспечивает защиту инвариантных областей TLP.

(продолжение)

Пакеты уровня транзакций несут признак одной из двух фаз транзакции – запрос (Request) и выполнение (Complete), последняя нужна не для всех типов транзакций.

Связь между запросами и выполнениями – по идентификатору транзакции (Transaction ID) из поля заголовка TLP.

Стандартный заголовок:



- TC – класс трафика
- TD – признак наличия дайджеста (CRC)
- EP – «отравленные» данные - признак ошибки данных и данные могут быть недействительными (poisoned data);
- Length – длина поля данных в dword

Качество обслуживания и виртуальные каналы

В PCI Express имеется поддержка дифференцированных классов по качеству обслуживания (QoS), обеспечивающая следующие возможности:

- выделять ресурсы соединения для потока каждого класса (виртуальные каналы);
- конфигурировать политику по QoS для каждого компонента;
- указывать QoS для каждого пакета;
- создавать изохронные соединения.

Для поддержки QoS применяется маркировка трафика: каждый пакет TLP имеет трехбитное поле метки класса трафика TC (Traffic Class). Это позволяет различать передаваемые данные по типам, создавать дифференцированные условия передачи трафика для разных классов. Порядок исполнения транзакций соблюдается в пределах одного класса, но не между разными классами.

Виртуальные каналы

Виртуальный канал VC (Virtual Channel) представляет собой физически обособленные наборы буферов и средств маршрутизации пакетов, которые загружаются только обработкой трафика своего виртуального канала.

На основе номеров виртуальных каналов и их приоритетов производится арбитраж при маршрутизации входящих пакетов.

Каждый порт, поддерживающий виртуальные каналы, выполняет отображение пакетов определенных классов на соответствующие виртуальные каналы. При этом на один канал может отображаться произвольное число классов.

По умолчанию весь трафик маркируется нулевым классом (TC0) и передается дежурным каналом (VC0). Виртуальные каналы вводятся по мере необходимости.

Поле «дайджеста»

Digest — 32-битный CRC-код. Длина всего пакета перечисленных полей кратна двойному слову (32 бит).

Признак «дайджеста» TD: единичное значение указывает на применение 32-битного CRC-кода в конце пакета, защищающего все поля пакета, не изменяемые в процессе его путешествия через коммутаторы PCI Express.

.

CRC –контроль (повторение)

CRC - Cyclical Redundancy Check - Контроль с помощью циклического избыточного кода. Способ контроля целостности данных при их передаче и хранении.

При помощи специального алгоритма вычисляется контрольная сумма пакета данных, которая передается вместе с ним. Алгоритм расчета контрольной суммы определяется используемым протоколом передачи данных.

Принимающее устройство повторно вычисляет контрольную сумму пакета данных.

Несовпадение рассчитанной и принятой контрольной суммы расценивается как ошибка передачи данных, при этом, как правило, принимающее устройство производит запрос повторной передачи ошибочного пакета.

Использование сообщений

Сообщения могут применяться для различных управляющих целей.

Эмуляция прерываний INTx# выполняется с помощью посылки сообщения с кодом установки либо снятия одного из 4 флагов прерываний (INTA-INTD).

Эмуляция PME#, а также других состояний энергопотребления, включая события недостатка питания, также выполняется с помощью сообщений.

Сообщения об ошибках передают один из трех кодов: исправимая (Correctable), не фатальная (Non-fatal) и фатальная (Fatal) ошибка.

Есть также сообщения о событиях Hot-plug (индикаторы Power и Attention, кнопка отключения и т.п.), а также событиях, определенных производителем.

Канальный уровень (Data Link Layer)

Отвечает за обеспечение целостности и достоверности данных, а также управление соединением.

На этом уровне пакеты уровня транзакций (TLP – Transaction Layer Packet) дополняются уникальным номером и контрольной суммой CRC.

Уровень проверяет порядок пакетов и контролирует их содержание, запрашивает пропущенные пакеты, сигнализирует о сбоях соединения, управляет состояниями соединения (неактивно, режим ожидания/инициализации, активно), служит для подачи сигналов энергопотребления, индикации ошибок и журналирования, обмена информацией управления потоком и т.д.

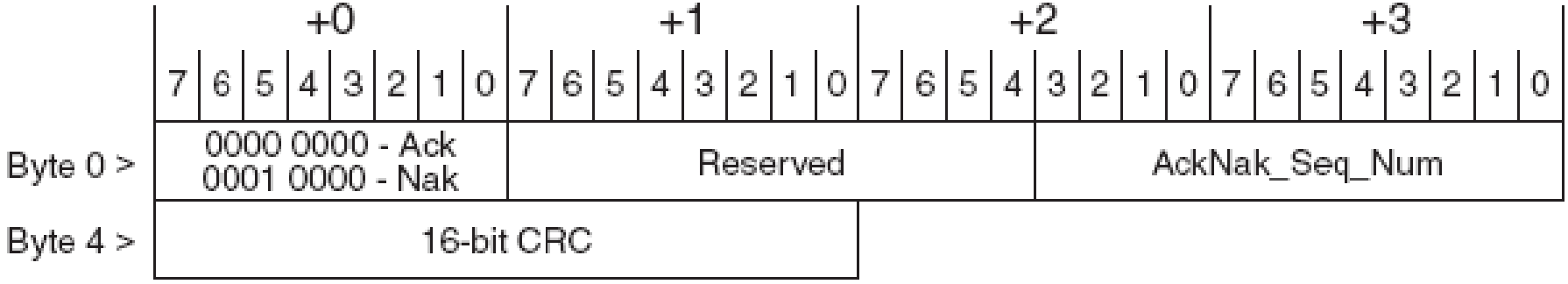
Пакеты DLLP

Специальные пакеты DLLP (Data Link Layer Packet) – служебные, данных не содержат, служат для управления соединением. Они не проходят через промежуточные узлы, распространяются только между портами.

Подразделяются на следующие типы:

- Ack – подтверждение прихода TLP с заданным номером
- Nack – запрос на повтор TLP с заданным номером
- Пакеты управления кредитами и VC
- Пакеты управления PM

DLLP содержит заголовок с типом пакета, информационное поле и 16-битный CRC (LCRC).



Качество обслуживания и виртуальные каналы

В PCI Express имеется поддержка дифференцированных классов по качеству обслуживания (QoS), обеспечивающая следующие возможности:

- выделять ресурсы соединения для потока каждого класса (виртуальные каналы);
- конфигурировать политику по QoS для каждого компонента;
- указывать QoS для каждого пакета;
- создавать изохронные соединения.

Для поддержки QoS применяется маркировка трафика: каждый пакет TLP имеет трехбитное поле метки класса трафика TC (Traffic Class). Это позволяет различать передаваемые данные по типам, создавать дифференцированные условия передачи трафика для разных классов. Порядок исполнения транзакций соблюдается в пределах одного класса, но не между разными классами

Виртуальный канал

Для дифференцирования условий передачи трафика разных классов в коммутирующих элементах PCI Express могут создаваться виртуальные каналы.

Виртуальный канал VC (Virtual Channel) представляет собой физически обособленные наборы буферов и средств маршрутизации пакетов, которые загружаются только обработкой трафика своего виртуального канала.

На основе номеров виртуальных каналов и их приоритетов производится арбитраж при маршрутизации входящих пакетов. Каждый порт, поддерживающий виртуальные каналы, выполняет отображение пакетов определенных классов на соответствующие виртуальные каналы. При этом на один канал может отображаться произвольное число классов. По умолчанию весь трафик маркируется нулевым классом (TC0) и передается дежурным каналом (VC0). Виртуальные каналы вводятся по мере необходимости.

Кредиты доверия

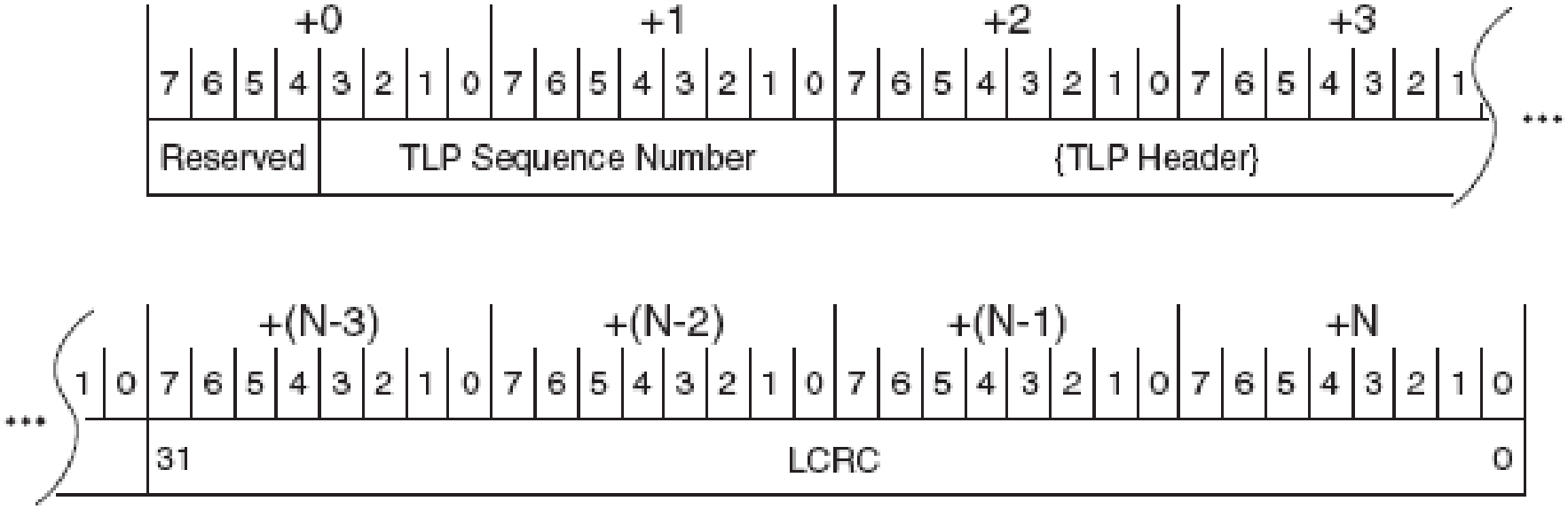
В протоколе PCIe для передачи пакетов по каналу связи используется механизм управления потоками на основе кредитов доверия. **Принимающее устройство выдает кредиты доверия на основе объема буферной памяти, имеющейся у принимающего устройства.** Передающему устройству запрещено начинать транзакции, для которых может потребоваться больше кредитов доверия, чем их заявило принимающее устройство. Для осуществления транзакции длина полезных данных, указываемая в заголовке запроса передающего устройства, должна точно совпадать с объемом передаваемых полезных данных и быть меньше или равна количеству кредитов доверия, имеющихся у принимающего устройства. Это может неоправданно ограничивать гибкость при передаче данных.

Оборачивание TLP

Уровень канала сопровождает пакет TLP уникальным номером и 32-битным кодом LCRC (Link CRC). TLP находится в retry-буфере до прихода DLLP типа Ack с тем же номером.

Код LCRC работает только в пределах одного соединения.

Существуют развитые правила запроса и выполнения повторов, таймеров ожидания ответа (в зависимости от размера пакета и ширины линии) и т.д.

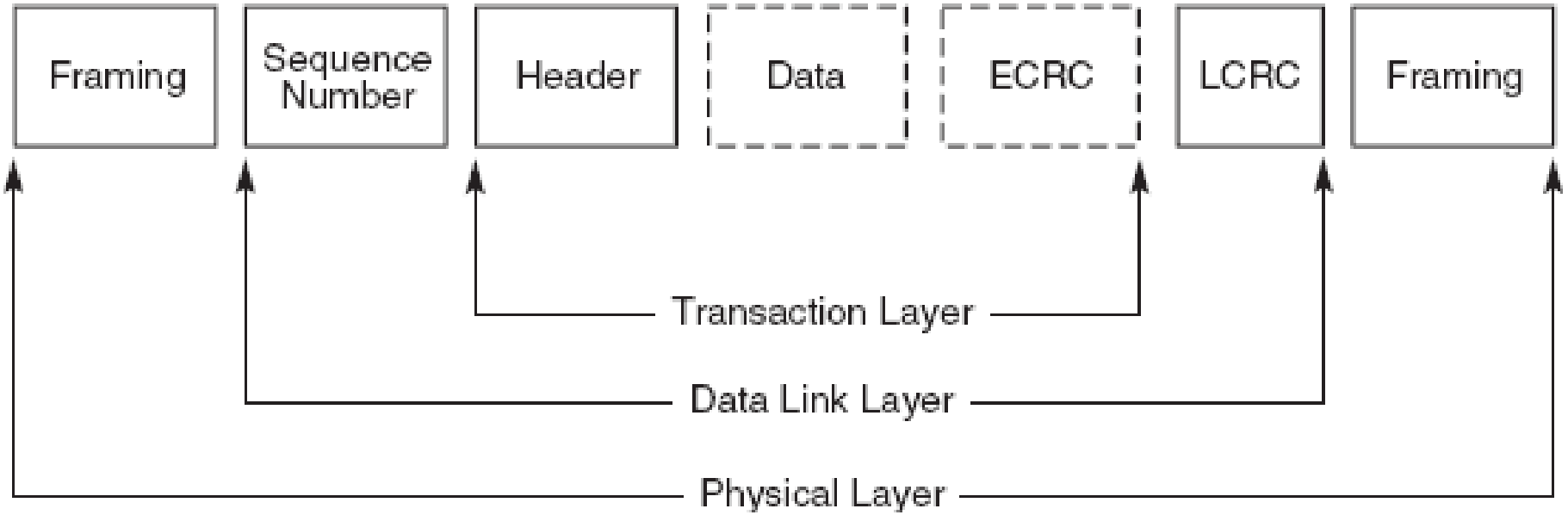


Физический уровень

Делится на два подуровня – логический и собственно электрический.

На логическом уровне байты полученных данных кодируются по схеме 8b/10b и преобразуются в 10-битные символы. Выполняется также скрэмблирование (если необходимо), распределение по линиям, кадрирование, обрамление служебными символами.

В результате данные принимают следующий вид:

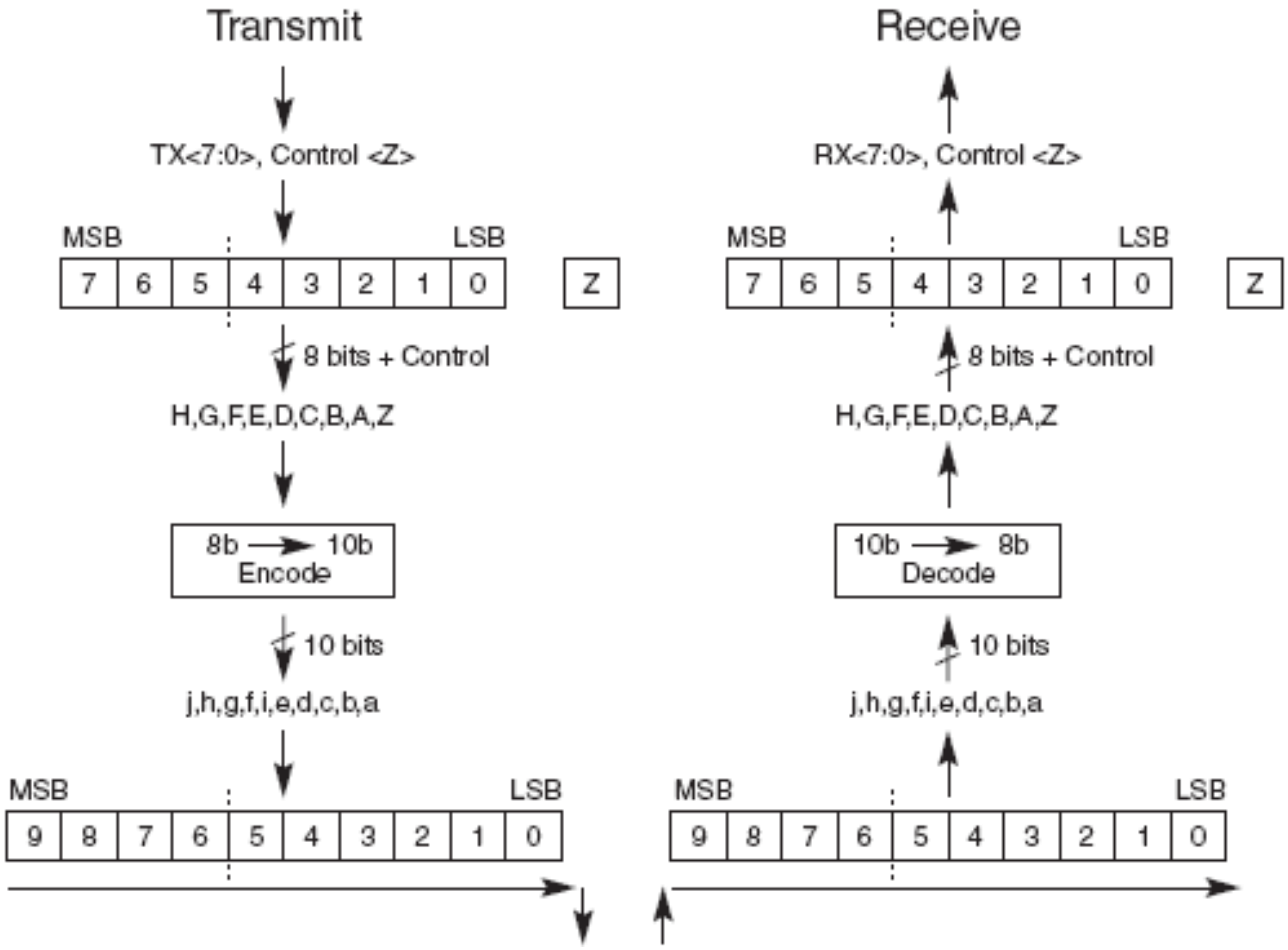


Кодирование 8b/10b

- В отличие от шин ISA, EISA и PCI, **в технологии PCI Express не предусмотрен тактовый генератор.** Устройства вправе начинать передачу в любой момент, как только им будет, что передавать. Такая свобода, с одной стороны, повышает быстродействие, с другой, порождает проблему. Предположим, что 1 кодируется напряжением +3 В, а 0 - напряжением 0 В. Если первые несколько байтов равны нулю, как получатель узнает о том, что ему передаются данные? Действительно - **последовательность нулевых битов трудно отличить от простоя канала. Эта проблема решается при помощи так называемой 8/10-разрядной кодировки.** Согласно этой схеме, 1 байт фактических данных кодируется при помощи 10-разрядного символа. Из 1024 возможных 10-разрядных символов выбираются такие, которые за счет достаточного количества фронтов без задающего генератора обеспечивают синхронизацию отправителя и получателя по границам битов. В силу применения 8/10-разрядной кодировки суммарная пропускная способность канала, равная 2,5 Гбайт/с, сужается до фактической пропускной способности 2 Гбайт/с.

Кодирование 8b/10b

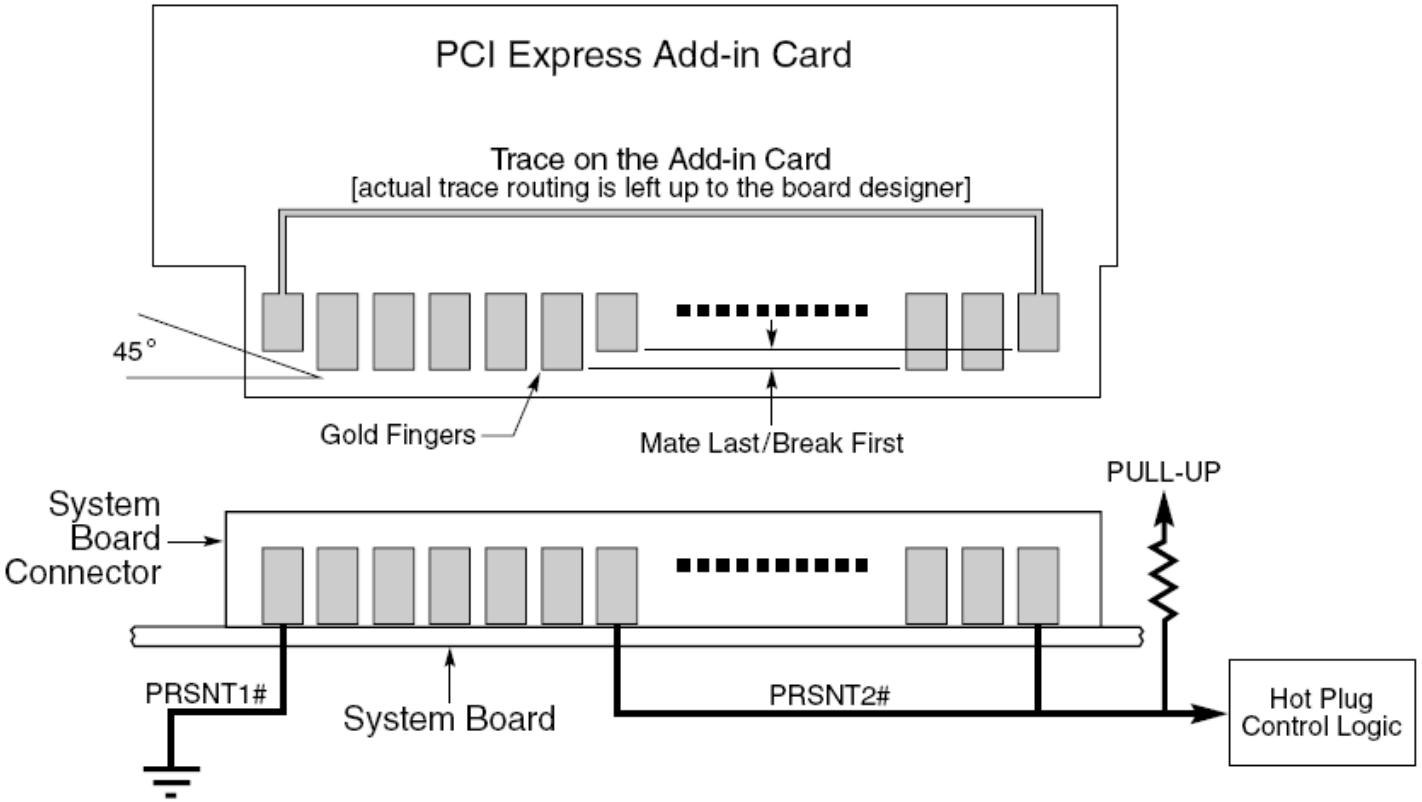
Кодирование 8b/10b выполняется по стандарту ANSI X3.230-1994 (или IEEE 802.3z). Младшие 5 бит отображаются на 6 бит, старшие 3 бита – на 4 бита, передаются младшим битом вперед



Карта PCI Express

По линиям PRSNT1/PRSNT2 производится определение наличия карты. Для каждого формата слота линия PRSNT2 находится в последнем ряду, PRSNT1 – в первом.

Подается питание +12 В, +3.3 В, +3.3Vaux. Также в слоте разведены интерфейсы SMBus и JTAG.



Карта PCI Express Mini Card

Специальный форм-фактор PCI Express Mini Card создан для карт расширения, устанавливаемых в мобильные компьютеры и мини-ПК.

Он предусматривает описание стандартных габаритов и разъема уменьшенного размера, а также дополнительных внешних выводов карты (антенна, светодиоды, сетевые розетки и т.д.).

Основное назначение карт Mini Card – сетевые и коммуникационные устройства (адаптеры WiFi, WiMax, Bluetooth, GPRS/CDMA/UMTS), которые должны быть модульными и легко заменяемыми.

Речь не идет о пригодности к замене самим пользователем. Проблема в другом: существующие законодательные нормы использования радиочастотного диапазона не позволяют использовать все типы сетевых устройств в некоторых странах. Производитель ноутбука должен выбирать тип коммуникационной карты в зависимости от страны назначения.

Карта Mini Card реализует два интерфейса – системный PCI Express x1 и периферийный USB

Карты ExpressCard

Организация PCMCIA, занимающаяся формализацией разработок в области карт расширения для ноутбуков с «горячим» подключением, предложила новый стандарт карт расширения – ExpressCard. От стандарта PC Card он унаследовал только некоторые из габаритов корпуса и общую конструкцию.

Фактически в корпусе модуля ExpressCard может быть помещено устройство с интерфейсом либо PCI Express x1, либо USB. В версии ExpressCard 2.0 обеспечена поддержка PCI Express 2.0 и USB 3.0, что позволяет устройствам получить канал с пропускной способностью 5 Гбит/с – достаточно для внешних винчестеров, ТВ-тюнеров, широкополосных модемов, виртуальных видеокарт и других требовательных устройств.

Функции управления энергопотреблением уже встроены в PCI Express и особенно USB, что сокращает стоимость внедрения ExpressCard.

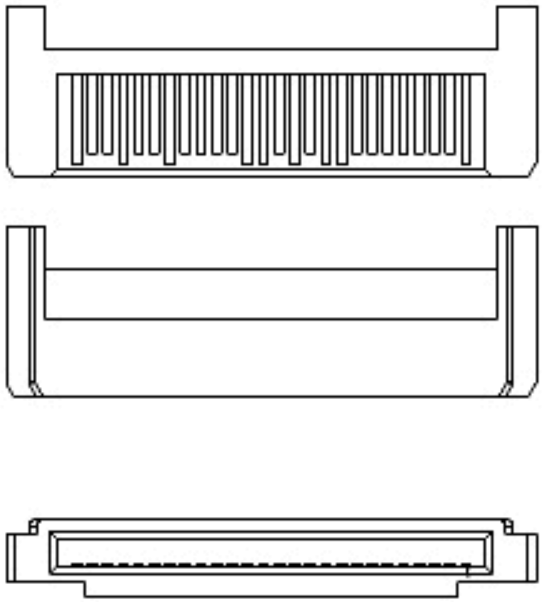
Физический интерфейс

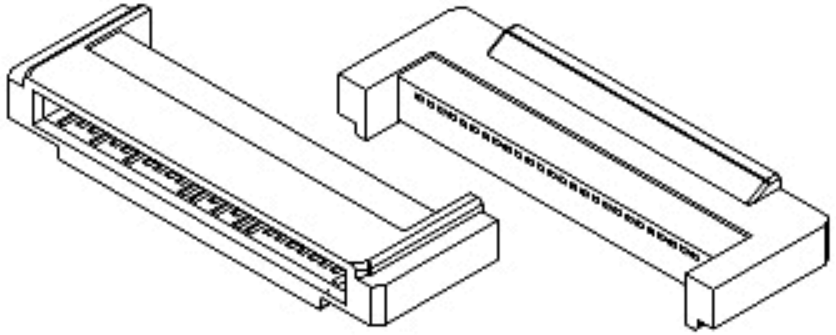
По сути ExpressCard описывает только физический интерфейс – размер модуля и формат разъемов.

Благодаря тому, что интерфейсы PCI Express и USB последовательные, удалось сократить размеры разъема (по сравнению с PC Card) и реализовать сразу два интерфейса.

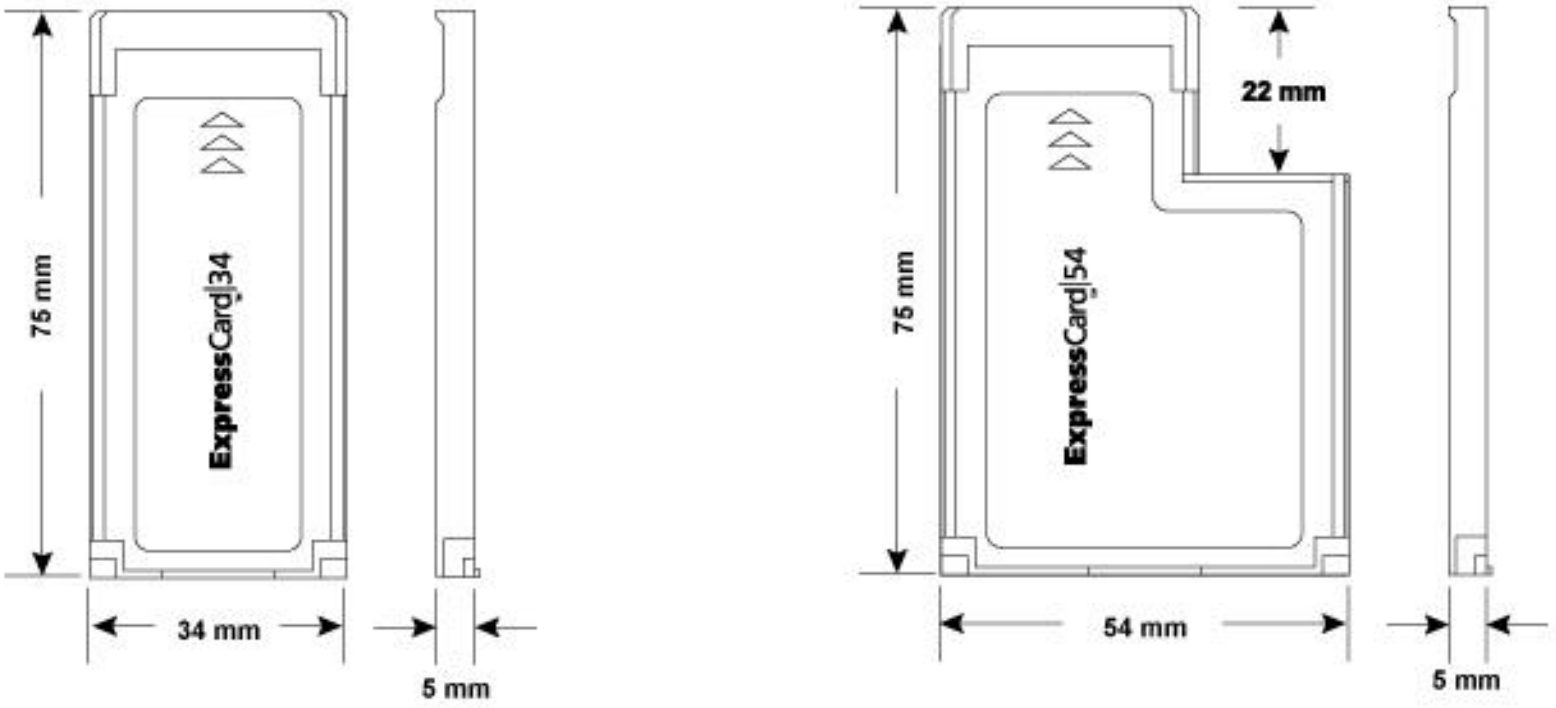
Карты ExpressCard имеют единую толщину (5 мм) и различаются только шириной – 34 мм или 54 мм (для устройств, которые не помещаются в корпус 34 мм), разъем идентичен. Слоты могут быть универсальными или только для устройств 34 мм.

Разъемы ExpressCard





Модули ExpressCard



Заключение

- возможность эффективно работать с различными структурами данных;
- - низкое энергопотребление и поддержку функций энергосбережения;
- - качество стратегий обслуживания;
- - поддержку "горячей замены" и "горячей установки" устройств;
- - обеспечение целостности данных и обнаружение ошибок на нескольких уровнях;
- - изохронную передачу данных;
- - узловую передачу при использовании чипов-мостов и одноранговую передачу с помощью коммутаторов;
- - многоуровневую технологию с поддержкой пакетной коммутации.