

## Раздел 2. Лекция 5.

Специализированные интерфейсы  
подключения накопителей. Часть 3

# Основные вопросы лекции

1. Интерфейс FC-AL.
2. Интерфейс SAS.
3. Протокол NVMe Express.

# 1. Fibre Channel Arbitrated Loop (FC-AL)

Вместо параллельного интерфейса SPI в жестких дисках может использоваться интерфейс FC-AL, который представляет собой физическую среду передачи данных с топологией «кольцо», *поверх которой работает SCSI*.

Отказ от коммутаторов позволяет существенно сократить стоимость без ущерба (для не нужной в данном случае) функциональности.

FC-AL представляет собой устаревший сетевой интерфейс, описанный 5-уровневой моделью. На нижнем уровне (FC-0) описан интерфейс типа «кольцо» (с трансляцией кадров от устройства к устройству) с применением либо оптоволоконных, либо электрический (коаксиальный) кабель. При этом физически могут присутствовать и концентраторы. Уровень FC-4 может занимать в том числе и SCSI.

FC-AL обеспечивает пропускную способность до 4 Гбит/с (400 Мб/с с учетом кодирования 8b/10b. Но главное преимущество – длина кабеля, которая для оптоволоконна может достигать 10 км.

# 1. Fibre Channel Arbitrated Loop (FC-AL)

В его основе лежит система команд SCSI-3. Это позволяет увеличить пропускную способность до 200MBps, FC-AL поддерживает до 126 самоконфигурирующихся устройств с возможностью «горячего» подсоединения без перезагрузки.

FC-AL - это спецификация оптоволоконных каналов связи с медным проводом. Такая конструкция имеет также повышенную помехоустойчивость, предоставляя некоторую избыточность доступа к каждому устройству.

- ❖ FC-AL самоконфигурируется и допускает оперативное (hot-plug) включение.
- ❖ Большинство хостов поддерживает FC-AL в режиме включения host-to-host.
- ❖ FC-AL может работать с коаксиальным кабелем длиной более 30m и с оптоволоконным до 10km.

# 1. Топология FC-AL

Порты FC могут быть соединены в конфигурации:

- «точка-точка»;
- «кольцо» (AL - Arbitrated Loop);
- посредством коммутатора (FC-Fabric).

Дуплексное соединение реализуется при помощи 2 противоположно направленных симплексных каналов, связывающих соответствующие передатчики и приемники.

Порты, связанные соединением «точка-точка», называются **N\_Port**, если же они работают в кольце, то **L\_Port**, а порты чипа Fabric - **F\_Port**.

Порт представляет собой микросхему, реализующую логику соединения на одном конце. Каждый порт содержит как минимум приемник и передатчик, а некоторые также и контроллер, и действует как репитер. Все порты функционируют в одинаковом режиме независимо от используемой топологии. Любая из FC-топологий опирается на соединение типа «порт-порт».

# 1. Иерархический стек протоколов: 5 уровней

- **FC-0.** Низший физический уровень, определяющий интерфейс связи со средой передачи. Он устанавливает требования к оптоволокну, коннекторам, оптическим и электрическим параметрам среды для различных скоростей передачи данных. Этот уровень обеспечивает физическую часть соединения "точка-точка".
- **FC-1.** Уровень, определяющий протокол передачи, включая алгоритмы преобразования параллельного кода в последовательный и обратно, а так же передачу спецсимволов и контроль ошибок.
- **FC-2.** Сигнальный протокол, определяющий транспортный механизм FC. В его функции входят формирование кадров из передаваемой последовательности байт, различные механизмы управления тремя классами услуг и средства диспетчеризации последовательности передаваемых данных.
- **FC-3.** Уровень, определяющий набор функций для повышения эффективности передачи, таких как распределение данных по нескольким портам, работающим параллельно для увеличения полосы пропускания канала, или доставка широковещательных сообщений множеству портов-получателей.
- **FC-4.** Высший уровень в сетке протоколов и определяет прикладной интерфейс FC. Он описывает правила, по которым происходит взаимодействие верхних уровней, использующих технологию FC.

# 1. Применение FC-AL

FC-AL в основном ориентирован на серверные приложения уровня high-end, где необходимо иметь быстрый доступ к ресурсам и где высокая помехоустойчивость - одно из важнейших преимуществ.

Сверхдлинные кабели также важны для этой области приложений, так как зеркальные диски обычно располагаются в разных зданиях для надежной защиты от непредвиденных ситуаций.

Спецификация FC-AL является естественным решением для создания на ее базе гигабитных опорных магистралей и локальных сетей, которые требуют широкой полосы пропускания для обработки изображений, видеоинформации и передачи сверхбольших объемов данных.

## 2. Концепция интерфейса SAS

Ввиду общей тенденции внедрения последовательных интерфейсов вместо параллельных назрела необходимость доработки и интерфейса SCSI.

В сфере, где он применяется, проблемы стоимости и удобства обслуживания, характерные для традиционного параллельного SCSI, не были очень острыми, однако дальнейшее наращивание быстродействия интерфейса было весьма затруднительным.

Интерфейс Serial Attached SCSI строился по той же концепции и на базе тех же аппаратных решений, что и Serial ATA, и у них очень много общего.

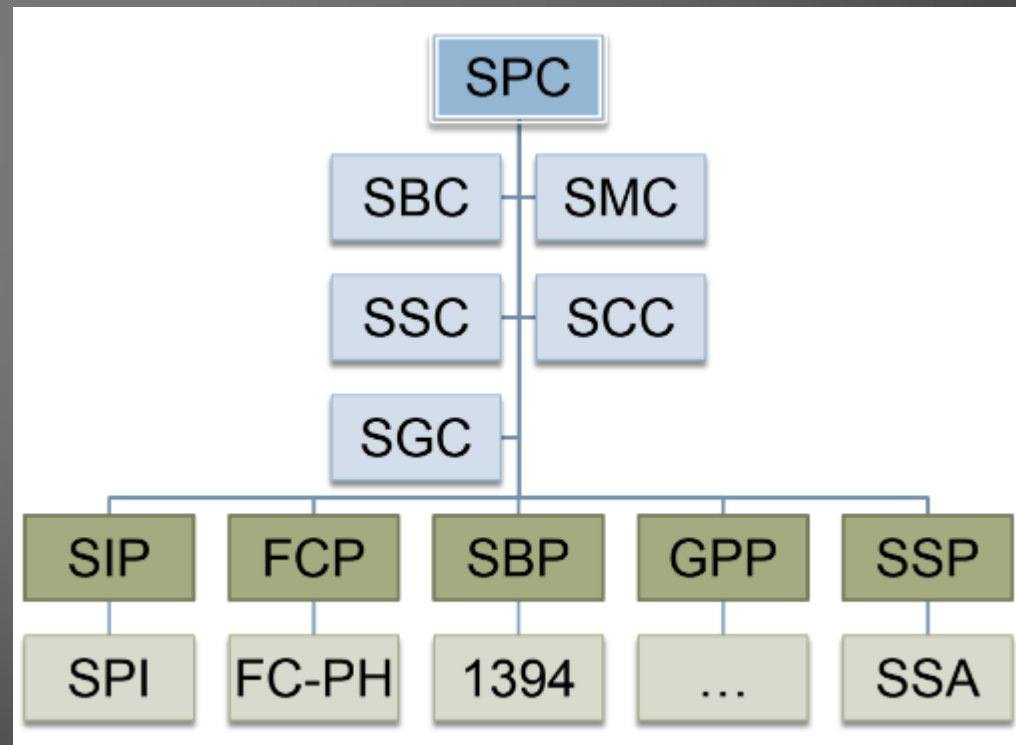
В частности, была обеспечена **полная совместимость контроллеров SAS с устройствами SATA** за счет унификации электрического и физического интерфейсов включения канального и транспортного слоев SATA в состав архитектуры SAS. Основная цель такого решения – достижение максимальной гибкости при формировании массива дисков.



## 2. Обобщенная архитектурная модель

Интерфейс SAS входит в общую архитектурную модель SCSI, предлагая иную среду передачи данных – последовательный интерфейс SSP.

- **SSP** (Serial Storage Protocol) – протокол последовательной памяти, реализуемый через архитектуру **SSA** (Serial Storage Architecture).



## 2. Архитектура SAS

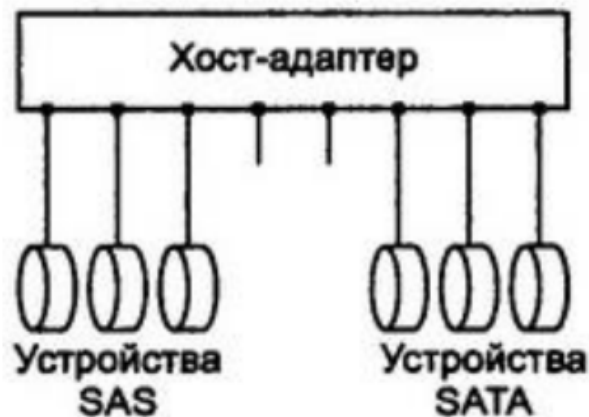
В то же время SAS имеет собственную архитектурную модель, в которой протокол SSP (Serial SCSI Protocol) является частью общей схемы и может быть заменен другими протоколами транспортного уровня.

Для физической передачи данных используется интерфейс, аналогичный Serial ATA. Однако общая концепция SAS намного сложнее, и протокол SATA *фактически является его подсистемой*.

SAS обеспечивает более сложную схему управления, обмена командами и данными, подключения и обслуживания, нежели SATA. Фактически это интерфейс нового поколения, использующий общие принципы SCSI (команды, адреса, ресурсы и т.п.).

## 2. Класс устройств

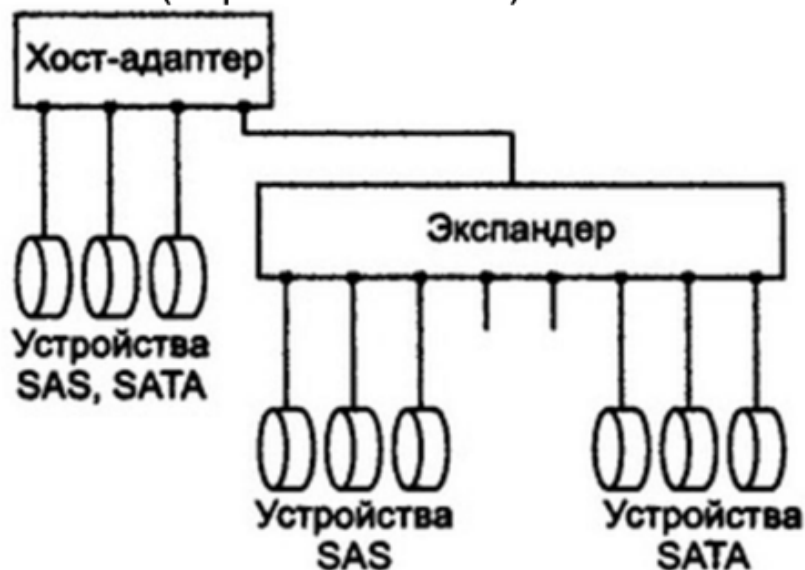
конечные устройства  
(end device)



Конечные устройства — это устройства SAS, являющиеся инициаторами или/и целевыми устройствами SCSI (или SATA).

Примеры:  
**SAS-диск,**  
**Оптический привод**

устройства-экспандеры  
(expander device).



Устройства-экспандеры служат для объединения конечных устройств в сложных конфигурациях - реализуют несколько дополнительных портов SAS для подключения к контроллеру.

Экспандер имеет собственный SAS-адрес для управления его функциями по протоколу SMP.

## 2. Хост-адаптер, экспандер

**Хост-адаптер**, содержащий 8 трансиверов, в котором возможны разные варианты конфигурации:

- один SAS-адрес на все трансиверы — это адаптер с одним, потенциально «очень широким» портом;
- два SAS-адреса, каждый на группу из четырех трансиверов — двухпортовый адаптер (удобен для подключения четырехканальными кабелями);
- восемь SAS-адресов — восьмипортовый адаптер, к которому можно подключить до 8 устройств SAS и SATA.

**Экспандер** может содержать и внутренние устройства SAS (со своими SAS-адресами), подключаемые к его портам с виртуальными трансиверами. Эти устройства могут, например, использоваться для управления блоком (питание, климат, защита и т. п.).

Каждый трансивер экспандера имеет собственный идентификатор (номер), уникальный в пределах экспандера. Порты экспандера могут служить для подключения инициаторов и целевых устройств SAS, а также других экспандеров. По этим портам будут передаваться кадры любых протоколов (SSP, STP, SMP). Экспандер может (не обязательно) содержать мосты STP/SATA (один или несколько), позволяющие к портам экспандера (узким) подключать устройства SATA. По этим портам будут передаваться только кадры SATA

## 2. Варианты последовательных протоколов SCSI

В разные годы предпринимались попытки применить тот или иной последовательный интерфейс для реализации SCSI, но они были или безуспешными, или не пригодными для подключения жестких дисков.

Interconnect	Standard	Year	Typical speed	Key features
Fibre Channel	FCP	1995	100 MB/sec	Optical
Serial Storage Architecture (SSA)	SSA-S2P, SSA-TL1, SSA-PL1	1996	20 MB/sec	IBM only vendor
Serial Storage Architecture (SSA)	SSA-S3P, SSA-TL2, SSA-PL2	1997	40 MB/sec	IBM only vendor
FireWire (IEEE 1394)	SBP-2	1998	50 MB/sec	
Fibre Channel	FCP-2	2002	200 MB/sec	
InfiniBand	SRP	2002	250 MB/sec	4x, 12x too
Ethernet	iSCSI	2003	~ 100 MB/sec	Gigabit Ethernet

## 2. Протоколы интерфейса SAS

Предусмотрен один физический интерфейс, но три различных протокола транспортного уровня:

- Serial SCSI Protocol (**SSP**), для серверных жестких дисков и других устройств хранения данных.
- Serial ATA Tunneling Protocol (**STP**), для жестких дисков SATA.
- Serial Management Protocol (**SMP**), для задач оптимизации и управления массивами жестких дисков.

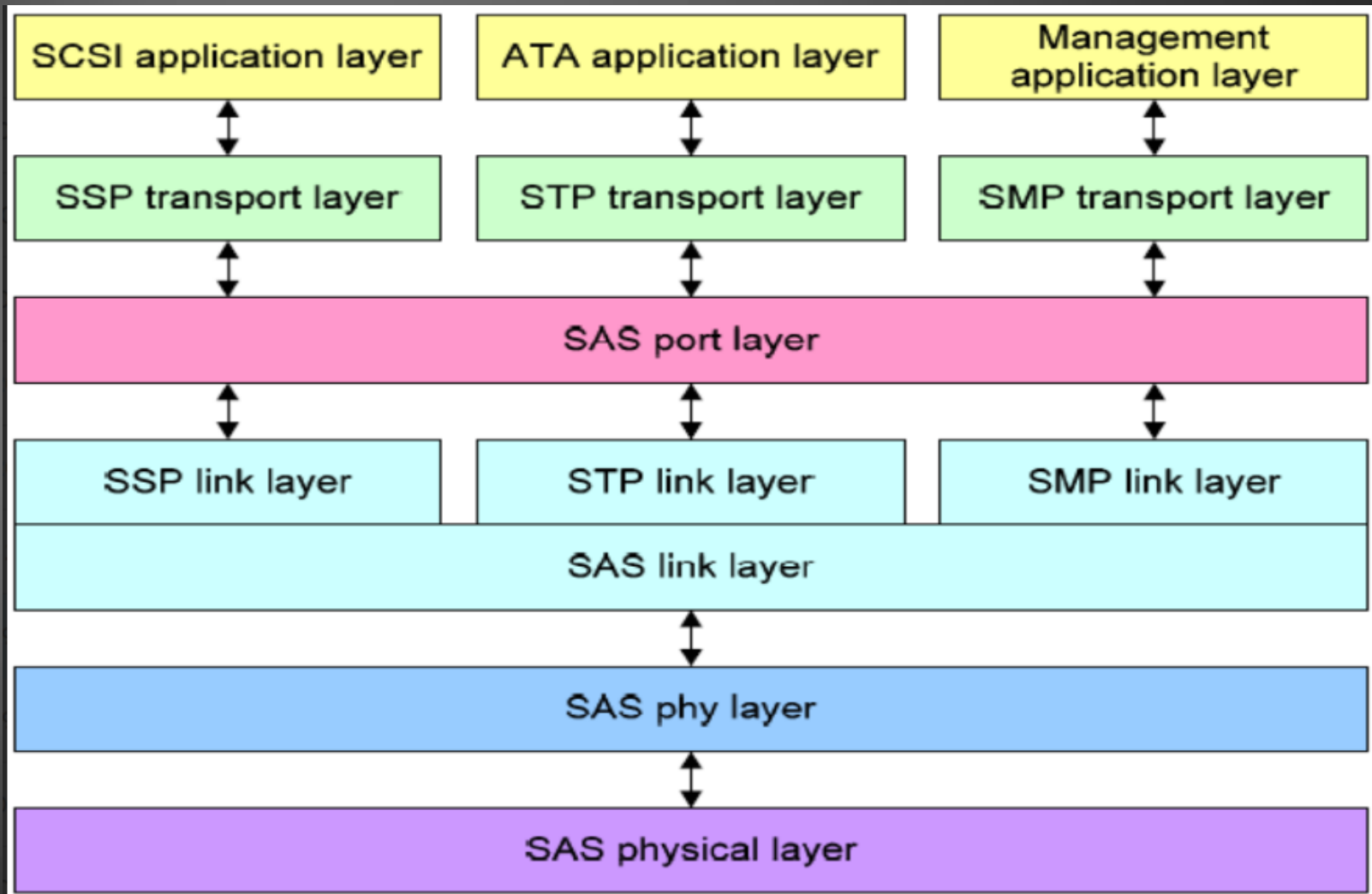
Поддержка SATA позволяет использовать один и тот же конструктив и набор компонентов для создания как экономичных, так и скоростных дисковых подсистем.

Использование SMP позволяет управлять экспандерами.

Протокол SAS содержит четыре традиционных уровня:

- физический (phy layer),
- коммуникационный (link layer),
- уровень портов (port layer)
- транспортный уровень (transport layer).

## 2. Архитектурная модель SAS



## 2. Уровни модели

- **Физический уровень** – определяет физические и электрические параметры приемников, передатчиков и кабелей;

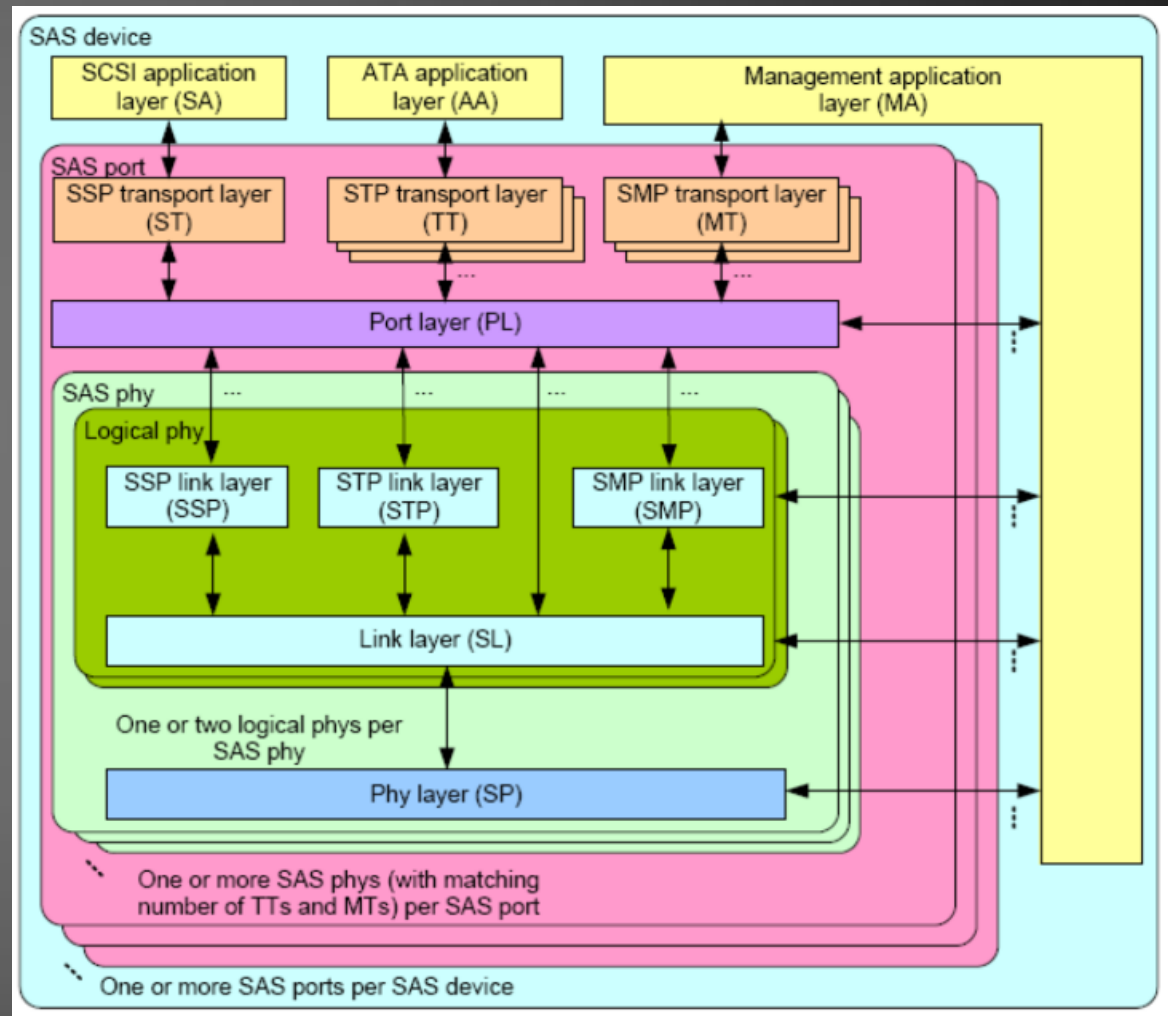
*Интерфейс физического уровня – определяет способ кодирования данных и специальную «внеполосную» сигнализацию для служебных целей.*

- **Канальный уровень** – обеспечивает идентификацию подключенных устройств и управления соединениями;
- **Транспортный уровень** – определяет структуры кадров и транспортные сервисы;
- **Прикладной уровень** – определяет процедуры выполнения команд SCSI в SSP и функции определения топологии и управления экспандерами в SMP.



## 2. Архитектура SAS

Объединение четырех уровней в каждом порте SAS означает, что программы и драйверы, используемые для работы с параллельными портами SCSI, могут с равным успехом использоваться и для обслуживания портов SAS, лишь с незначительной модификацией.



## 2. Уровни портов SAS

Уровни приложений, включающие драйверы и собственно приложения, создают специфические задания для транспортного уровня, который, в свою очередь, инкапсулирует команды, данные, статусы и пр. в SAS-фреймы и перепоручает их передачу *уровню портов*.

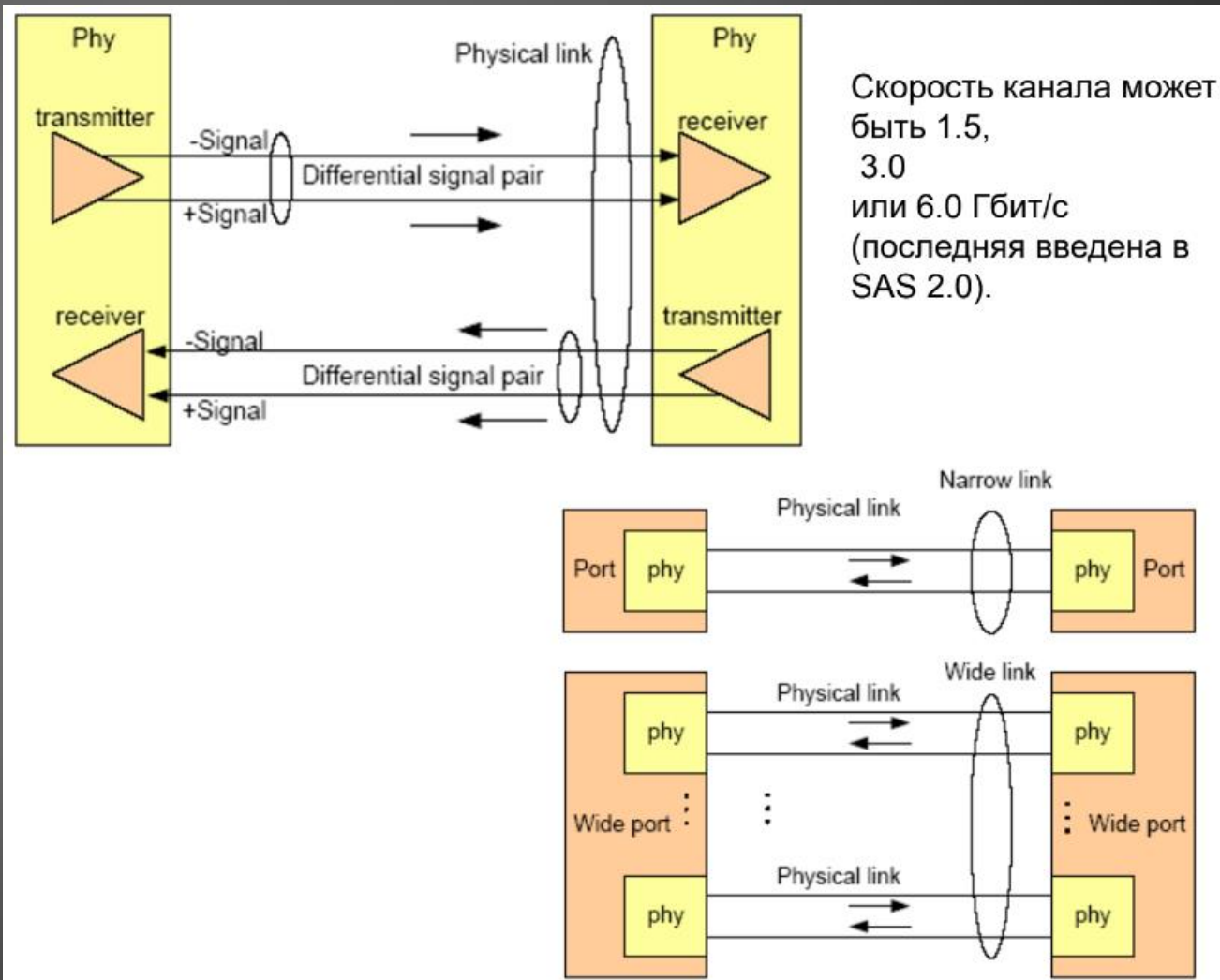
Разумеется, транспортный уровень также отвечает за прием SAS-фреймов и с уровня портов, дизассемблирование принятых фреймов и передачу контента уровню приложений.

**Уровень портов SAS отвечает за:**

- обмен пакетами данных с коммуникационным уровнем (link layer) в порядке установления соединений;
- за выбор физического уровня, с помощью которого будет осуществляться передача пакетов одновременно на несколько устройств.

Под физическим уровнем SAS подразумевается соответствующее аппаратное окружение - трансиверы и модули кодирования, которые подключаются к физическому интерфейсу SAS и отправляют сигналы по проводным цепям.

## 2. Уровни портов SAS



## 2. Физический уровень

Уровень Physical Layer описывает механические и электрические характеристики кабелей и разъемов, соединяющих устройства (жесткие диски) с экспандерами и хост-адаптерами.

Тип кабеля для подключения одного устройства унаследован от SATA, разъем несколько отличается, но является совместимым – содержит контакты второго порта с обратной стороны между разъемами питания и интерфейса.

Второй порт при подключении жесткого диска при помощи кабеля не задействуется, однако при подключении встык через backplane (панель оснастки) второй порт позволяет удвоить пропускную способность интерфейса.

Разъем питания может быть стандартным (SATA) либо специальным (SAS), с передачей сигнала активности диска по одному из резервных контактов.

## 2. Связь уровней Phy и Port

Для того, чтобы обеспечить увеличение пропускной способности путем удвоения (учетверения и т.д.) количества физических каналов, были введены дополнительные уровни.

В терминах SAS:

- Phy – аппаратный блок обработки данных, имеющий уникальный адрес SAS и поддерживающий работу канального уровня.
- Физические Phy соединены кабельным интерфейсом, аналогичным Serial ATA (две дифференциальные пары, по одной на передачу и прием).
- Логические Phy, входящие в состав физических Phy, могут поддерживать один из трех протоколов – SSP, STP или SMP.
- Порт SAS может соответствовать одному Phy (узкий порт, narrow port), либо объединять несколько (широкий порт, wide port).
- Если у нескольких Phy один адрес SAS, то они не могут разделяться на несколько портов.

## 2. Физический и канальный уровень

Рассмотрим подробнее интерфейс физического уровня и канального уровня.

На уровне физического интерфейса определяется:

- логическое кодирование 8B/10B;
- «внеполосная» сигнализация • согласование скоростей.

Все передачи организуются в виде двойных слов (32 бита).

Внеполосная сигнализация используется для обнаружения факта подключения устройства, определения его типа (SAS, SATA) и согласования скорости интерфейса.

На канальном уровне обеспечивается идентификация устройства, т.е. определение типа устройства (конечное, экспандер) и адреса порта, производится согласование скоростей путем добавления в передаваемый поток специальных заполнителей, не несущих информативности.

Канальный уровень так же обеспечивает управление соединением.

## 2. Доставка данных

Для открытия соединения инициатор посылает запрос либо непосредственно целевому устройству, если оно подключено непосредственно к инициатору, либо экспандеру, через который это целевое устройство подключено. Целевое устройство либо подтверждает, либо отклоняет запрос на соединение, но если между ними находится экспандер, то как только к нему поступил запрос, он пересылает его далее, а обратно посылает специальный сигнал, информирующий инициатора о том, что запрос находится в процессе арбитража.

Для закрытия соединения необходимо, чтобы приемник и передатчик переслали друг другу сообщения об окончании передачи, после чего оба устройства передают сигнал о закрытии соединения.

В случае, если к одному порту устройства приходит одновременно несколько запросов на соединение, то преимущество отдается запросу с наибольшим временем ожидания. Используя такой вариант доступа к разделяемому ресурсу, обеспечивается справедливость арбитража и ограниченная задержка обслуживания.

## 2. Доставка данных

Для гарантированной доставки данных, передаваемых по интерфейсу SAS, для каждого кадра данных, полученных целевым устройством, **генерируется квитанция** о принятии, либо непринятии кадра.

Инициатор, после отправки кадра, ожидает квитанции в течении 1 мс, если в течении этого времени пришла квитанция о принятии, то кадр считается успешно отправленным, иначе, если пришла квитанция о непринятии, либо квитанция не пришла в течении 1 мс, то кадр считается не переданным и влечет за собой закрытие соединения.

**Передача данных является не блокирующей, т.е. передатчик не ожидает приема квитанции об успешной доставке предыдущего кадра, а сразу же передает следующий.**

Управление потоком передачи основано на **кредитах**. Для того, чтобы передать кадр данных, передатчик должен иметь ненулевой кредит.

Кредит передатчику выдает приемник, при этом приемник должен иметь возможность принять полноразмерный кадр данных. Значение кредита может достигать 255.



## 2. Доставка данных

Отдельным вопросом стоит работа с устройствами SATA, т.к. интерфейс SATA не предусматривает наличие соединения. В следствие этого возможны два варианта работы:

- если устройство SATA непосредственно подключено к хост-адаптеру – тогда обмен данными происходит полностью по интерфейсу SATA;
- если же устройство SATA подключено к хост-адаптеру через экспандер, то работа происходит по протоколу STP, который открывает соединение, затрачивая на это дополнительное время, но в дальнейшем работа происходит напрямую, используя команды интерфейса SATA.

## 2. Экспандер

Экспандер не только физически коммутирует устройства, но и занимается маршрутизацией пакетов между Phy, подключенными к различным его портам.

Устройства могут обмениваться данными через порты экспандера, минуя хост-адаптер.

Маршрутизация может выполняться одним из трех способов:

- Прямая – по SAS-адресу Phy.
- Табличная – по таблице соответствия SAS-адресов физическим Phy.
- Субтрактивная – для каналов восходящих (к хост-адаптеру или экспандеру).

Экспандер может содержать встроенное SAS-устройство, доступное для управления.

## 2. Соединения

Ввиду наличия проблемы маршрутизации обмен пакетами выполняется в виде сессий соединения двух устройств с различными адресами SAS.

Соединение начинается с отправки кадра OPEN с SAS-адресом приемника. Приемник, получив кадр, отвечает примитивом OPEN\_ACCEPTED. Примитив CLOSE завершает соединение.

Широкий порт может устанавливать несколько соединений одновременно (к одному или разным портам, по числу линков этого порта).

У жестких дисков имеется два узких порта, а у хост-контроллера могут быть порты широкие, для работы с несколькими устройствами одновременно.

Ширина конкретного соединения может быть более одного Phy.

## 2. Согласование скоростей

Поскольку соединение может проводиться по разным сегментам, которые в общем случае обладают различной пропускной способностью (1.5, 3 или 6 Гбит/с), необходимо соблюдать единую скорость потока, равную пропускной способности самого «узкого» места в маршруте.

Достигается это за счет вставки «пустых» символов (1 или 2).

## 2. Протокол SSP

Данные передаются кадрами, которые начинаются с примитива SOF и заканчиваются примитивом EOF.

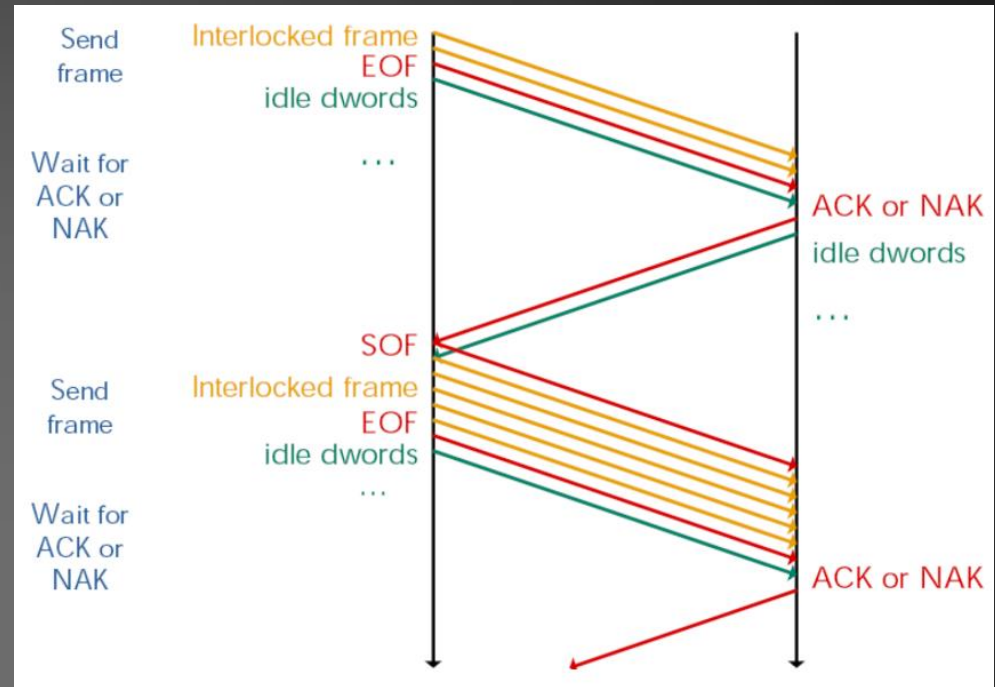
В ответ на каждый кадр должен прийти примитив ACK или NACK.

Кадры посылаются в обе стороны одновременно.

Разрешение на посылку кадра вторая сторона высылает в виде кредита – примитива RRDY.

В режиме SATA кадры передаются только в одну сторону, формат примитивов и заголовков блоков данных отличен от SAS. Кроме того, нет понятия «соединение», т.к. данные идут по маршруту «точка-точка».

## 2. Транспортный протокол SAS

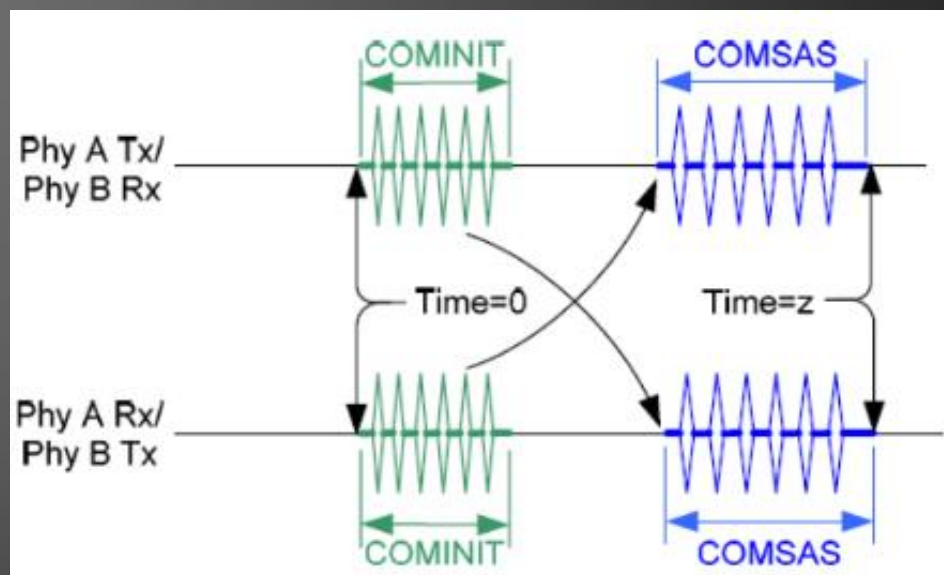


Транспортный протокол SAS (Serial Attached SCSI) SSP (Serial SCSI Protocol) является дуплексным.

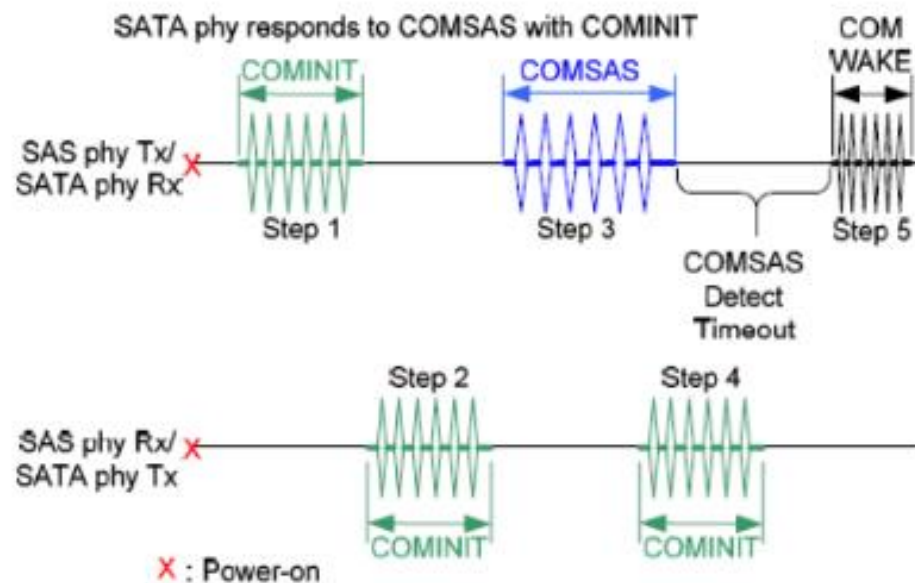
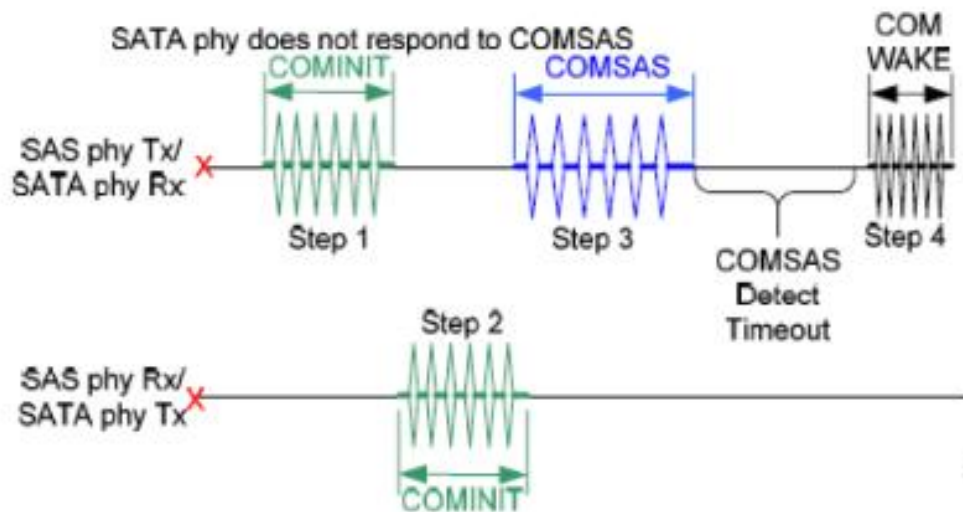
- Перед обменом требуется установление соединения (SATA не требует установления соединения).
- Кадр состоит из примитива SOF (Start Of Frame), двойных слов с данными и EOF.
- Ответом на каждый кадр является примитив ACK или NAK.
- Управление потоком на основе кредитов (как в FC), примитив RRDY.
- SSP кадр состоит (аналогично FC) из заголовка (24 байта), данных (до 1024 байт), заполнителя (от 0 до 2 байт), CRC (4 байта).

## 2. Уровень Phy. Режимы SAS и SATA

- Каждый физический канал может работать в режиме SAS или SATA, а если это восходящий канал экспандера – то в каждом по очереди.
- Выбор режима осуществляется в момент подключения по сигналам **OOB (Out of Band)**, которые представляют собой чередование всплесков (**примитивы Align**) и пауз ( $A+ = A- = 0$ ) с различной частотой.
- Если после OOB COMINIT последует COMSAS, то канал будет работать как SAS, иначе – только как SATA.



## 2. Уровень Phy. Режимы SAS и SATA





## 2. Транспортный уровень

Описывается формат кадра, порядок обмена ими. Формат кадра позаимствован у протокола FC-AL.

Всего предусмотрено 5 видов кадров:

- Command;
- Data;
- XFER\_READY;
- Response;
- Task.

Byte	Field(s)		
0	Frame Type		
1 to 3	Hashed Destination SAS address		
4	Reserved		
5 to 7	Hashed Source SAS address		
8 to 9	Reserved		
10	Reserved	Retransmit	Rsvd
11	Reserved	Number of Fill Bytes	
12 to 15	Reserved		
16 to 17	Tag		
18 to 19	Target Port Transfer Tag		
20 to 23	Data Offset		
24 to m	Information Unit		
m to (n-3)	Fill bytes, if needed		
(n-3) to n	CRC		

## 2. Сравнение SAS и SCSI

- ❖ Интерфейс **SAS** использует последовательный протокол передачи данных между несколькими устройствами - использует меньшее количество сигнальных линий.
- ❖ Интерфейс **SCSI** использует общую шину. Таким образом, все устройства подключены к одной шине, и с контроллером одновременно может работать только одно устройство. Интерфейс SAS использует соединения точка-точка — каждое устройство соединено с контроллером выделенным каналом.
- ❖ В отличие от SCSI, **SAS** не нуждается в терминации шины пользователем. В SCSI имеется проблема, связанная с тем, что время распространения сигнала по разным линиям, составляющим параллельный интерфейс, может отличаться. Интерфейс SAS лишён этого недостатка.
- ❖ **SAS** поддерживает большое количество устройств (> 16384), в то время как интерфейс SCSI поддерживает 8, 16, или 32 устройства на шине.
- ❖ **SAS** обеспечивает более высокую пропускную способность (1.5, 3.0, 6.0 или 12.0 Гбит/с). Такая пропускная способность может быть обеспечена на каждом соединении инициатор-целевое устройство, в то время как на шине SCSI пропускная способность шины разделена между всеми подключёнными к ней устройствами. контроллеры SAS могут поддерживать подключение устройств с интерфейсом SATA, при прямом подключении - с использованием протокола SATA, при подключении через SAS-экспандеры - с использованием туннелирования через протокол STP (SATA Tunneled Protocol).
- ❖ **SAS**, также как и параллельный SCSI, использует команды SCSI для управления и обмена данными с целевыми устройствами.

## 2. Сравнение SAS и SATA

- ❖ SATA-устройства идентифицируются номером порта контроллера интерфейса SATA, в то время как устройства SAS идентифицируются их WWN-идентификаторами (WWN — *World Wide Name*). Для подключения SATA-устройства к домену SAS используется специальный протокол STP (*Serial ATA Tunneled Protocol*), описывающий согласование идентификаторов SAS и SATA.
- ❖ Устройства SATA 1 и SAS поддерживают *тегированные очереди команд* TCQ (*Tagged Command Queuing*). В то же время, устройства SATA 2 поддерживают как TCQ, так и NCQ (*Native Command Queuing*).
- ❖ SATA использует набор команд ATA, который позволяет работать с жёсткими дисками, в то время как SAS поддерживает более широкий набор устройств, в том числе жёсткие диски, сканеры, принтеры и др. (Накопители на оптическом диске, подключаемые через SATA, на самом деле являются целевыми устройствами SCSI, для доставки SCSI команд к которым используется SATA).
- ❖ Аппаратура SAS поддерживает связь инициатора с целевыми устройствами по нескольким независимым линиям: в зависимости от реализации можно повысить отказоустойчивость системы и/или увеличить скорость передачи данных. Интерфейс SATA версии 1 такой возможности не имеет. В то же время, интерфейс SATA версии 2 использует дубликаторы портов для повышения отказоустойчивости.
- ❖ Преимущество SATA — в низком энергопотреблении и невысокой стоимости оборудования, а интерфейса SAS — большей надёжности.

### 3. Протокол NVMe Express

NVMe Express (NVMe) — это масштабируемый интерфейс хост-контроллера. Разработан для корпоративных и клиентских систем. NVMe подключает твердотельные диски (SSD) по шине PCI Express.

Интерфейс позволяет использовать все преимущества накопителей на энергонезависимой памяти. NVMe Express применяется в вычислительных комплексах любого масштаба — от мобильных устройств до центров обработки данных.



### 3. История NVMe Express

SSD существенно отличаются от HDD скоростными характеристиками и внутренними процессами ввода/вывода. «Дисковые» протоколы SATA и SAS не позволяют использовать весь потенциал SSD, поэтому было решено создать «с нуля» протокол, учитывающий особенности твердотельной памяти.

На Intel Developer Forum в 2007 году был представлен NVMeHCI (Non-Volatile Memory Host Controller Interface) — прототип будущего интерфейса.

Спецификации NVMe разрабатываются рабочей группой «NVMe Express Workgroup», в которую входит более 90 компаний. Первая версия 1.0 вышла в свет в марте 2011 года. В 2019 году готова спецификация NVMe Express 1.4.

### 3. Преимущества NVMe Express

Скорость	Масштабируемость	Эффективность
<ul style="list-style-type: none"><li>— Учтены особенные временные характеристики SSD</li><li>— Исключен дополнительный контроллер</li><li>— Параллельная передача данных по нескольким линиям PCI (локальное подключение)</li><li>— Сокращенный набор команд</li><li>— Параллельная обработка запросов множеством ядер</li></ul>	<ul style="list-style-type: none"><li>— Скорость передачи будет расти с повышением скорости линии PCI</li><li>— Параллельная обработка запросов множеством ядер позволяет наращивать производительность системы хранения</li><li>— NVMe-oF предоставляет ресурсы удаленным серверам обработки</li></ul>	<ul style="list-style-type: none"><li>— Используются все преимущества современных твердотельных накопителей</li><li>— Раздельное использование дорогостоящего оборудования. Поддерживаются механизмы виртуализации и разделяемого использования</li><li>— Сокращение энергозатрат</li></ul>

### 3. Преимущества NVMe Express

Разработчики интерфейса комплексно подошли к проектированию эффективного перспективного интерфейса с учетом современных тенденций в отрасли.

Главная задача — соответствие скоростных характеристик интерфейса NVMe уровню твердотельных накопителей SSD решена блестяще. Помимо этого, в интерфейс заложены механизмы, позволяющие строить эффективные масштабируемые вычислительные системы.

Множество путей доступа и разделяемая логическая структура (shared namespace) позволяют строить системы высокой доступности и коллективно использовать высокоскоростные диски.

Поддержка виртуализации SR-IOV упрощает использование накопителей в виртуальных средах.

### 3. Преимущества NVMe Express

Благодаря механизмам сетевого расширения интерфейса NVMe-oF, разнесенные внутри датацентра и даже географически разнесенные серверы получают доступ к подсистеме NVMe SSD.

В то же время, на пути между потребителем данных — процессором и поставщиком — накопителем по возможности удалены все промежуточные звенья.

Современный процессор содержит PCIe контроллер внутри себя, уже несколько поколений — это не отдельный чип управления. Процессоры Intel Scalable вкупе с драйвером UEFI получили встроенный хост-контроллер NVMe.



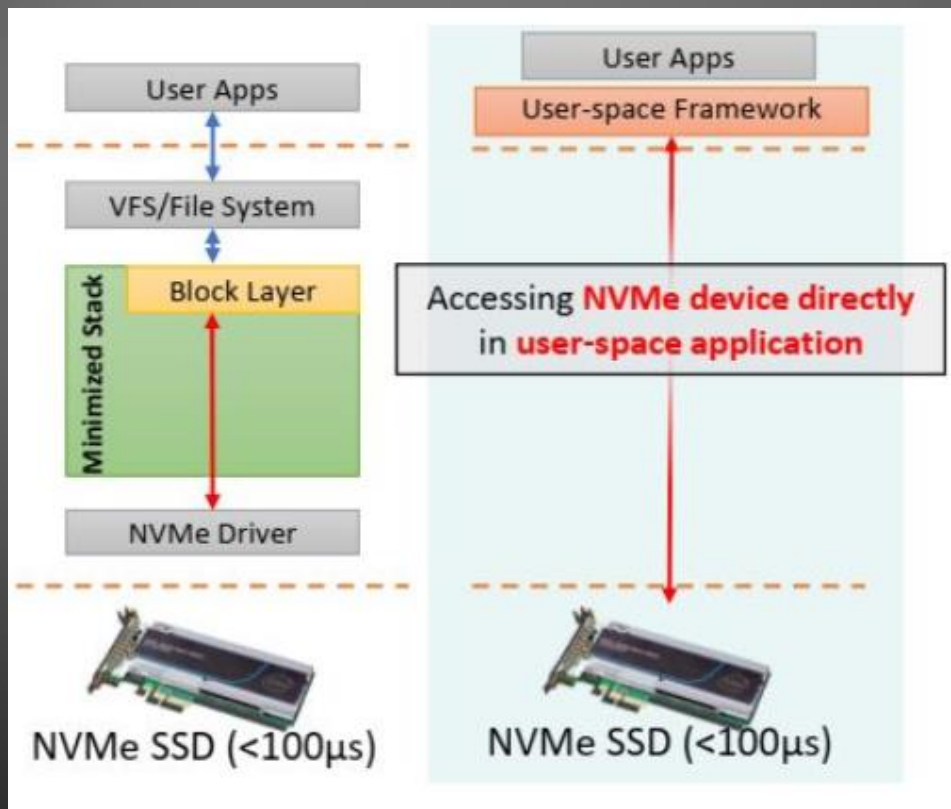
### 3. NVMe Zoned Namespaces

NVMe Zoned Namespaces (ZNS) ратифицирована NVM Express в июне 2020 как дополнение спецификации 1.4a.

Зонирование пространства имен придает SSD свойства следующего этапа эволюции накопителей. Более тонкое управление ресурсами драйва позволяет повысить его производительность, увеличить эффективность его использования.

### 3. NVMeDirect Framework

Параллельно с интерфейсом NVMe развивается новый, можно сказать — революционный механизм доступа к данным. NVMeDirect Framework позволяет приложению обращаться к данным, расположенным в хранилище, минуя операционную систему.



«Прямой доступ к накопителю»  
может принести еще порядка  
10% к производительности