Технический вебинар: NVMe.

Есть ли задачи для нового интерфейса? Анонс новых накопителей от HGST

Григорий Никонов, системный инженер

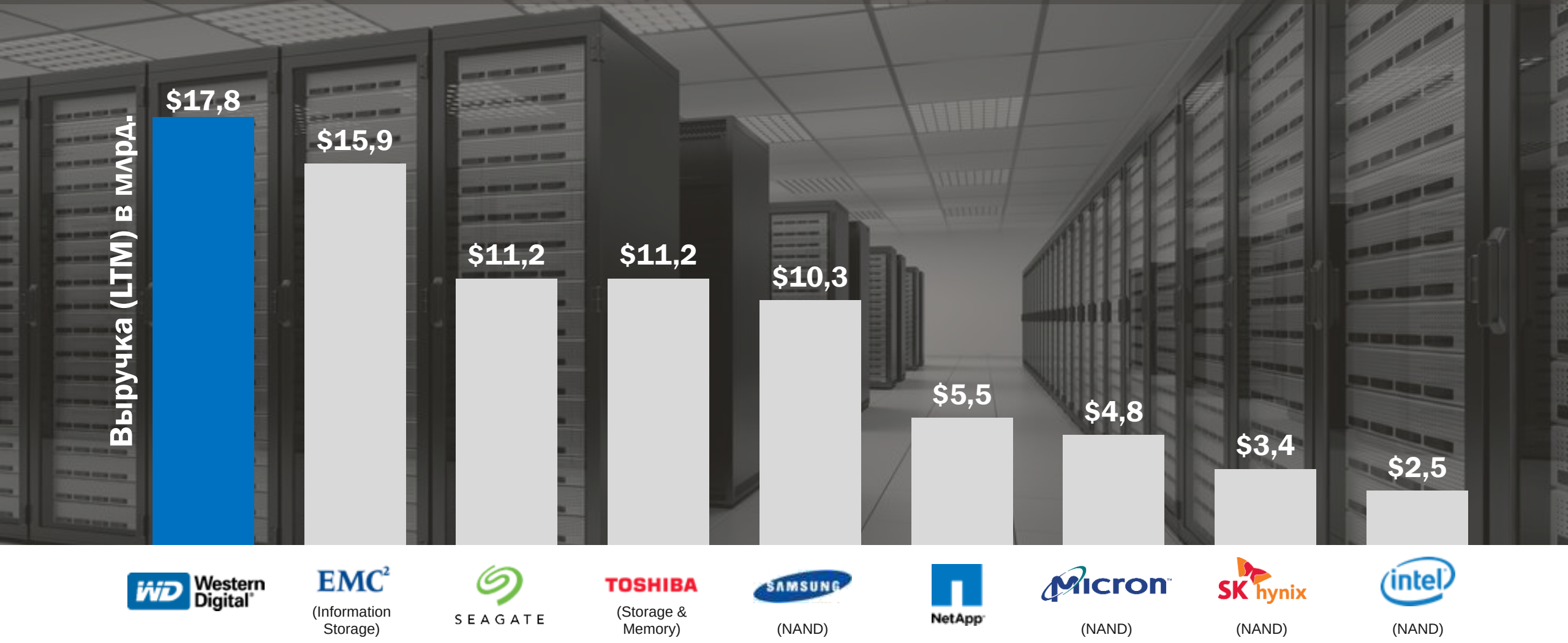
grigory.nikonov@hgst.com


HGST
a Western Digital brand


WD
a Western Digital brand

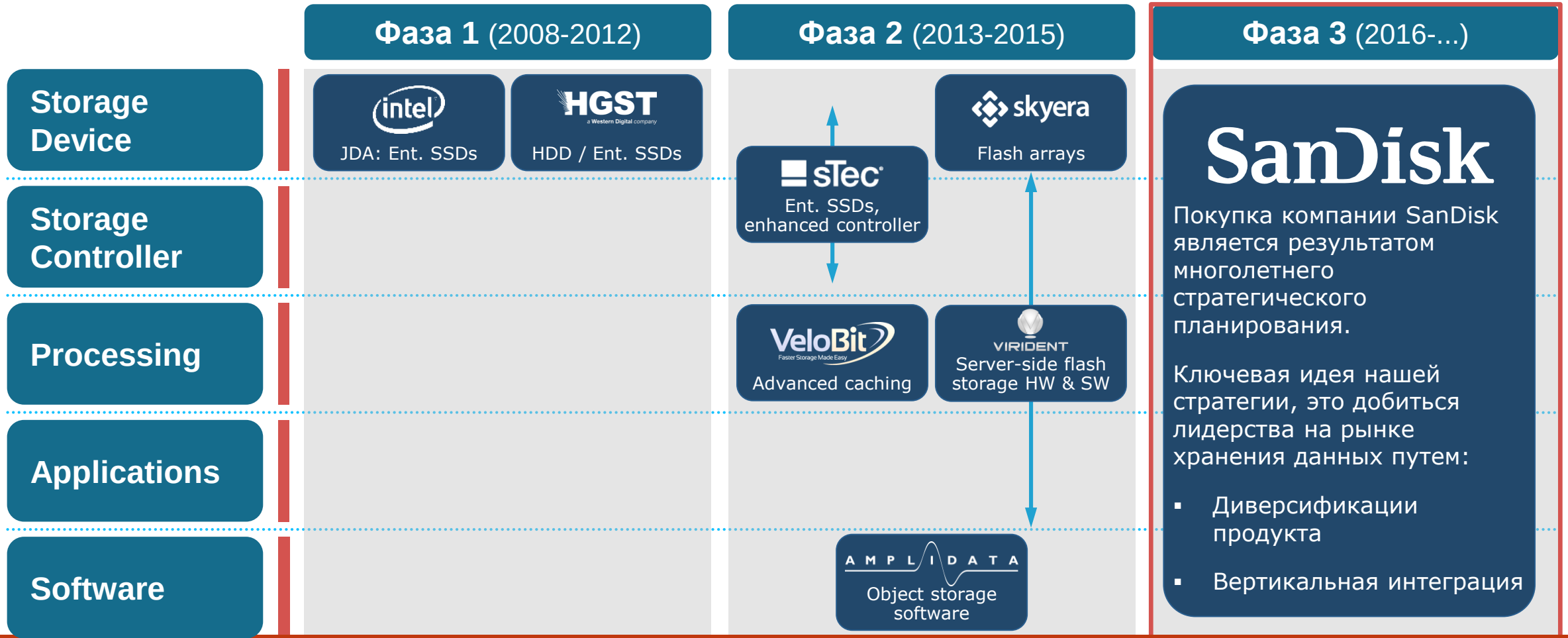

SanDisk®
a Western Digital brand

Индустрия хранения данных по состоянию на 1 июля 2016.

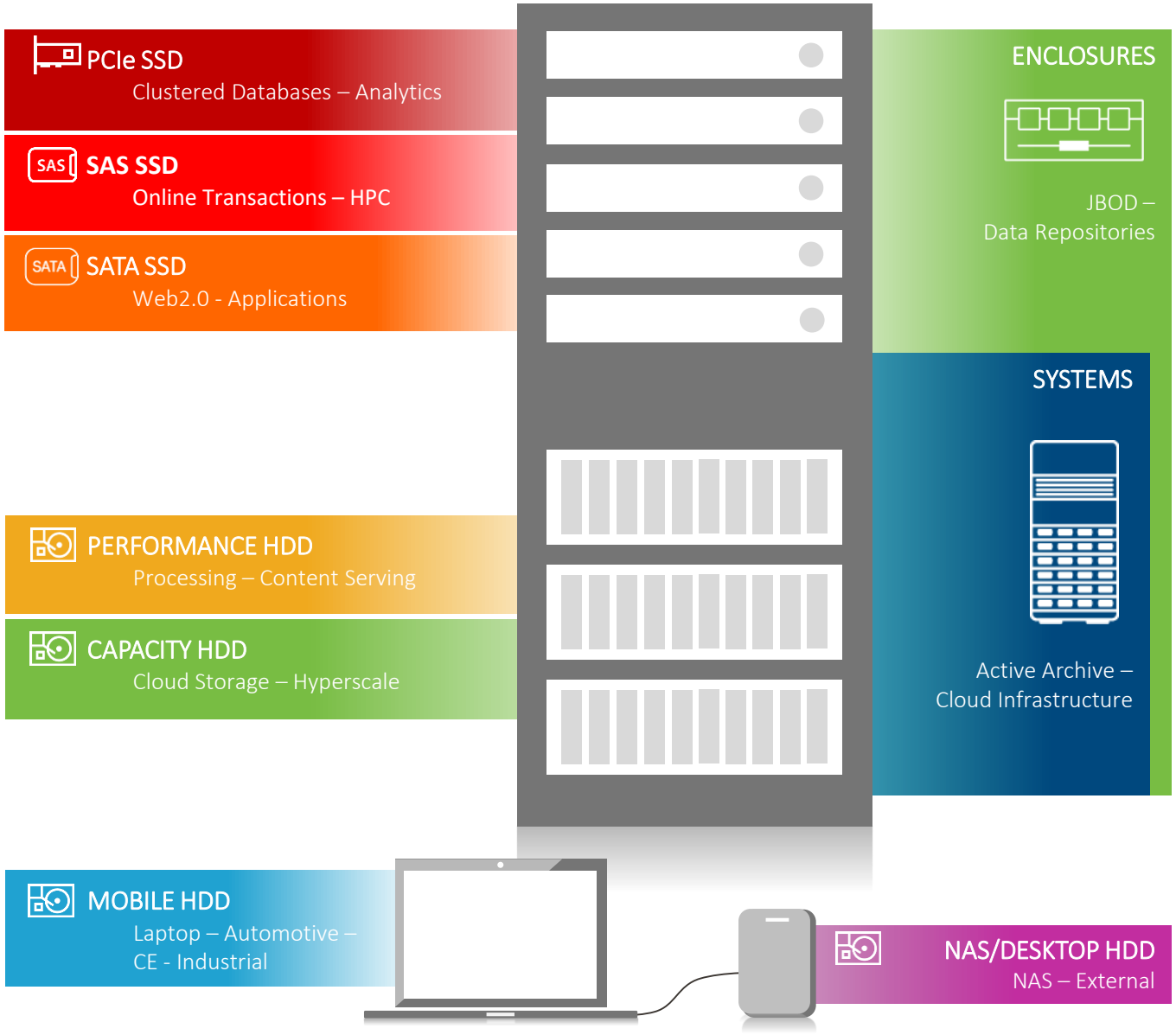


Наши стратегические приоритеты

Прошлые приобритения



Расширение наших возможностей и инвестирование для постоянного роста



В 1956 году мы придумали первый в мире жесткий диск

На этом мы не остановились

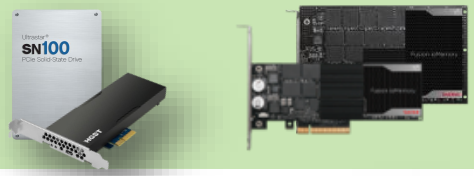
Сегодня HGST предлагает инновации во всех сегментах продуктов для хранения данных – от супер-быстрых твердотельных накопителей до архивных систем с максимальной плотностью размещения в мире.

HGST/SNDK для любых задач

Write Intensive

DWPD > 10

Caching
In-Memory Compute



Mix Use

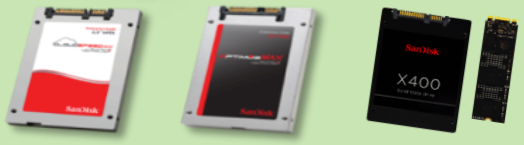
2 < DWPD < 3

Search
OLTP
VDI & Virtualization

Read Intensive

0.5 < DWPD < 1

Exchange Server
Analytics
Image Retrieval
Web Search
VOD/Media Streaming



Very Read Intensive

DWPD < 0.5

Data Warehouse
Web Server Logs
Content Repository

Commercial

Laptops Desktops
Commercial
Embedded
Server Boot

История SATA и SAS

- 1984/1994 IDE -> ATA-1 (Master/Slave)
- 1996 ATA-2 (WD EIDE + STX FastATA, LBA, Block transfer)
- 1997 ATA-3 (SMART)
- 1998 ATAPI (SCSI over ATA - CD, DVD, tape)
- 2003 SATA (AHCI, NCQ)

Hardware	SATA	SAS	PCIe
Software	AHCI	SCSI	NVMe

- 1978/1981 SCSI-1
- 1994 SCSI-2 (Fast)
- 2003 SCSI-3 (Ultra)
- 2005 SAS-1 (3 Gbps)
- 2009 SAS-2 (6 Gbps)
- 2013 SAS-3 (12 Gbps)
- 2017 SAS-4 (24 Gbps) ?

Появление NVMe

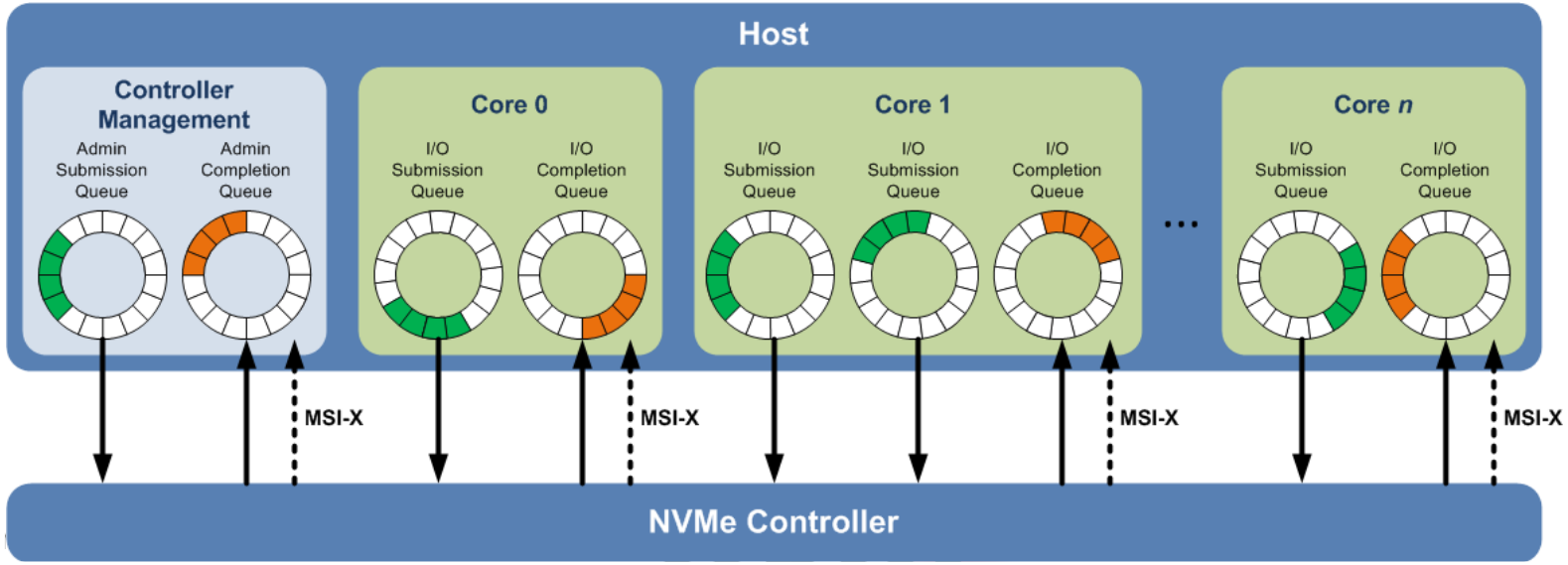
- 2011 - Появление Promoter Group, NVMe v 1.0
- 2012 - NVMe 1.1
- 2014 - NVMe 1.2
- 2014 - Образование компании NVM Express Organization

- Стандартизация регистров, функционала и набора команд
- Изначально создан для NAND и NVM следующего поколения
- Спроектирован для корпоративного и клиентского использования

PROMOTERS GROUP

Технические основы NVMe

- Глубокие очереди 64K x 64K
  - Рекомендовано 32-128 Ent / 2-8 Client
- Компактный набор команд - 13 обязательных
  - ATA: 100+ legacy commands / SCSI: 250+
- Поддержка MSI-X
  - Снижение нагрузки на CPU, устранение узких мест
- Работа с любым видом NVM
  - Совместимость с NAND и будущими типами памяти
- Эффективная работа на 4K
  - Все параметры команды в одном запросе 64B
- Опциональный функционал для клиентских и корпоративных дисков



Эволюция NVMe устройств

2015

Performance AIC

Эволюция NVMe устройств

2015
Performance AIC

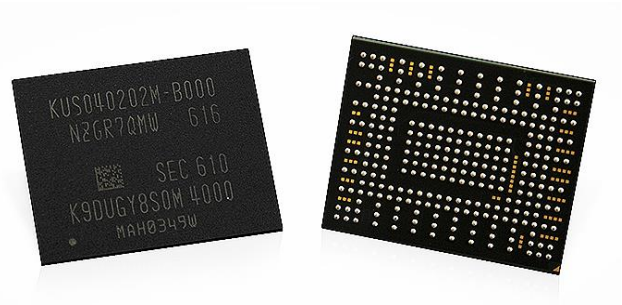
2016
Performance AIC & U.2
Enthusiast AIC
M.2 consumer

Эволюция NVMe устройств

2015
Performance AIC

2016
Performance AIC & U.2
Enthusiast AIC
M.2 consumer

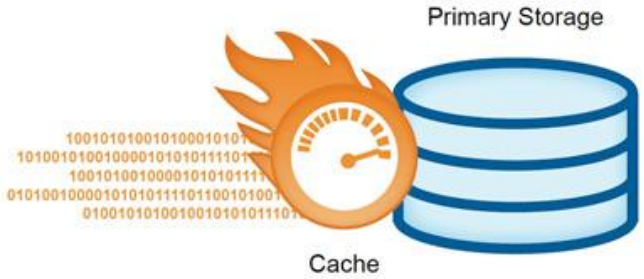
2017
Performance AIC & U.2
2x2 Performance U.2
Essential U.2
Enthusiast AIC
M.2 consumer
BGA



Задачи для NVMe

2015

Performance AIC



Microsoft SQL Server

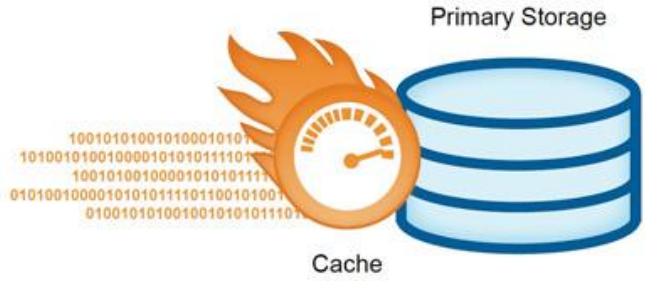
ORACLE

MySQL

PostgreSQL

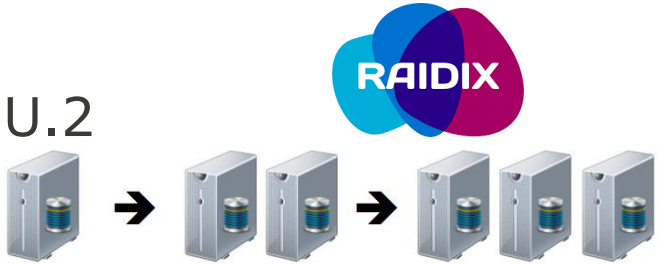
Задачи для NVMe

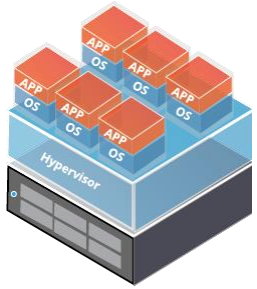
2015
Performance AIC



Microsoft SQL Server
ORACLE
MySQL
PostgreSQL

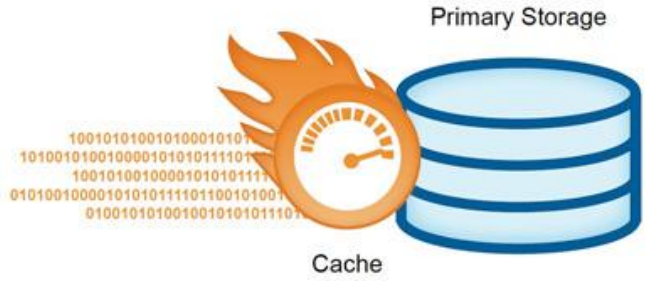
2016
Performance AIC & U.2
Enthusiast AIC
M.2 consumer





Задачи для NVMe

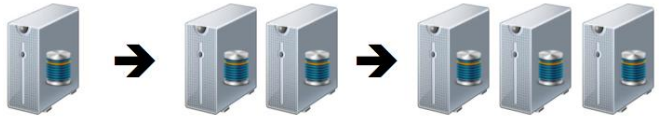
2015
Performance AIC

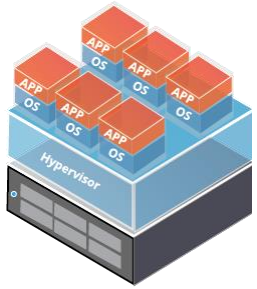


Microsoft SQL Server
ORACLE
MySQL
PostgreSQL

2016
Performance AIC & U.2
Enthusiast AIC
M.2 consumer

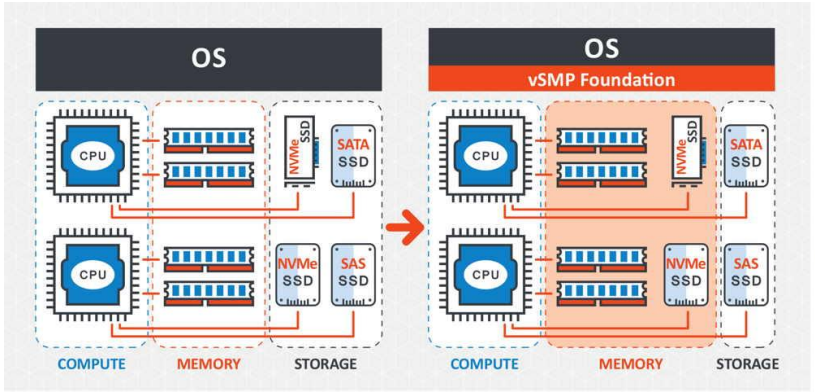
RAIDIX



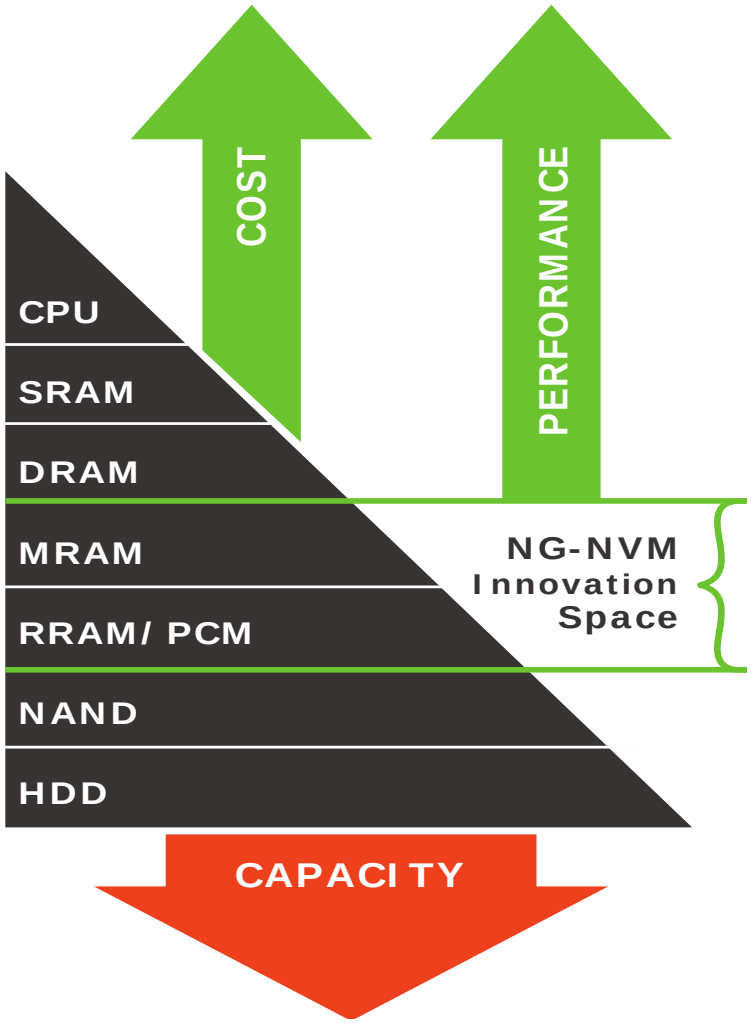


There is no cloud
it's just someone else's computer

2017
Performance AIC & U.2
2x2 Performance U.2
Essential U.2
Enthusiast AIC
M.2 consumer
BGA

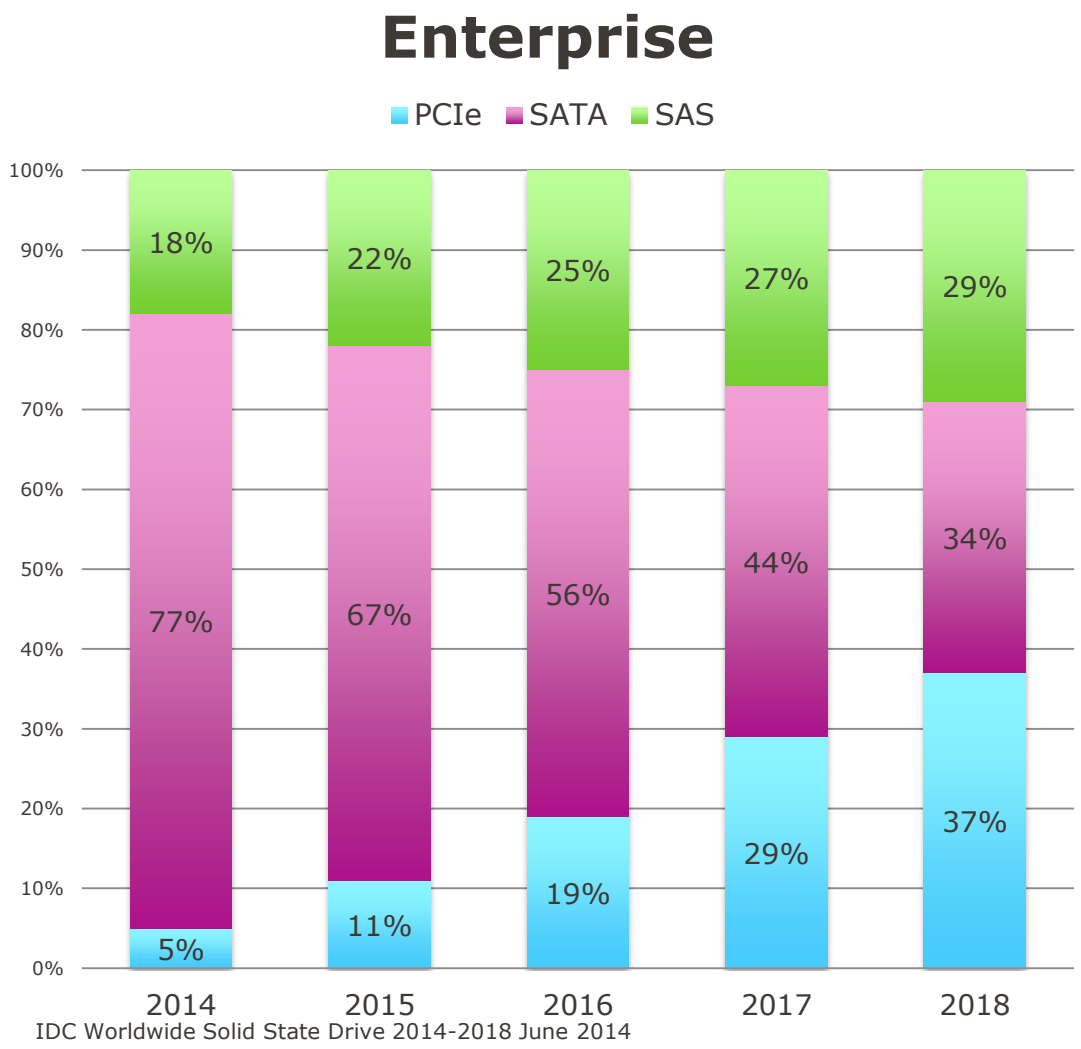


18+

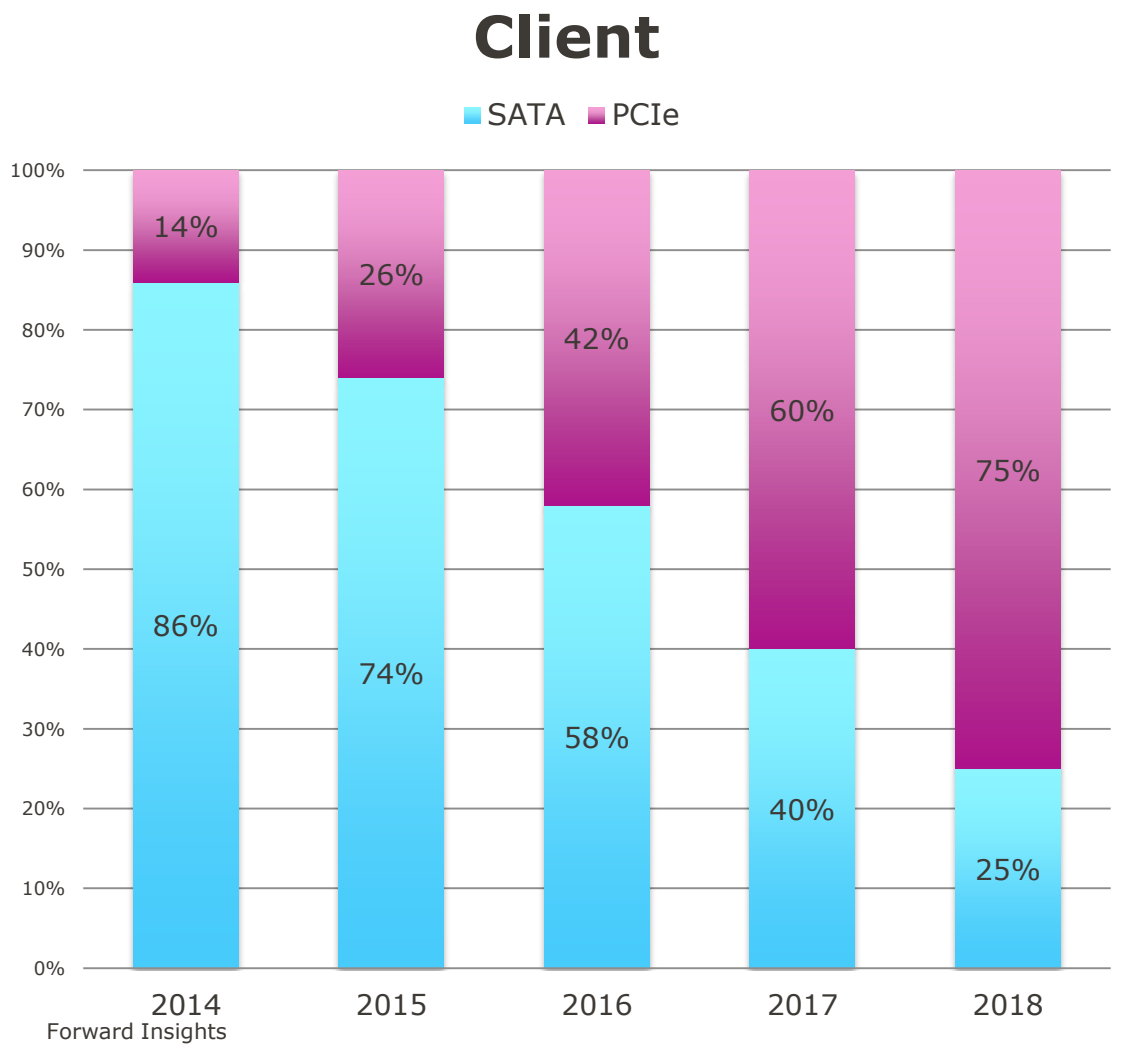

COST
PERFORMANCE
CPU
SRAM
DRAM
MRAM
RRAM/ PCM
NAND
HDD
NG-NVM
Innovation
Space
CAPACITY


PCIe SSD
Расширение DRAM
Замена DRAM
Универсальная память
2015
2016
2017
2018
2019
2020
2021

Наступление NVMe



IDC Worldwide Solid State Drive 2014-2018 June 2014



Forward Insights

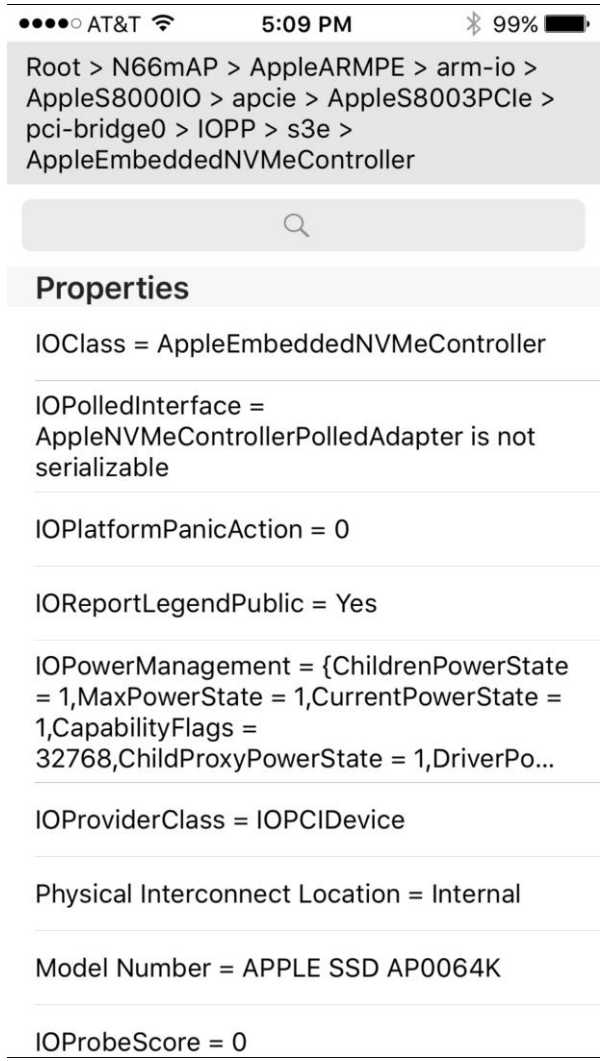
Знаете ли вы?.. iPhone 6s использует NVMe

Root > N66mAP > AppleARMPE > arm-io > AppleS8000IO > apcie > AppleS8003PCIe > pci-bridge0 > IOPP > s3e > AppleEmbeddedNVMeController

Properties

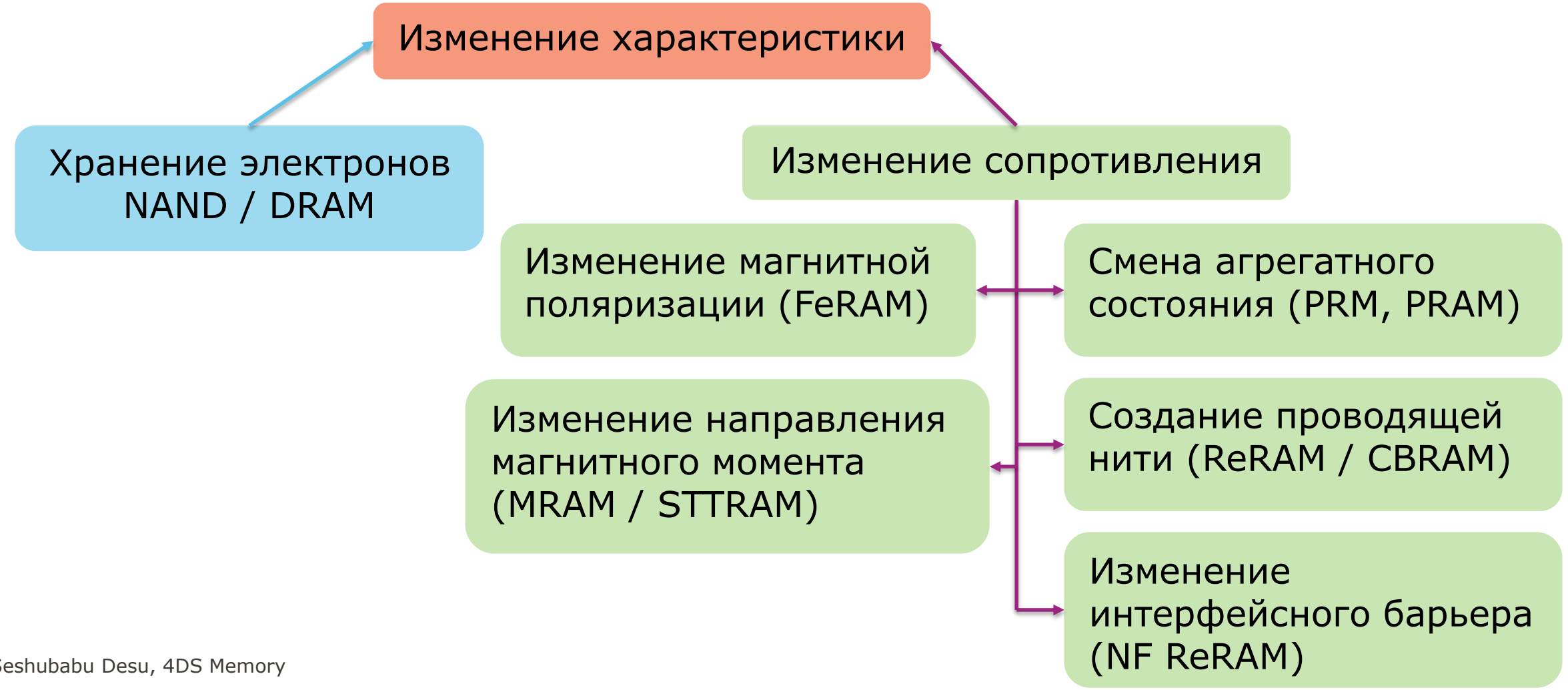
Physical Interconnect = PCI-Express

Controller Characteristics = {cell-type = 3,controller-unique-id = 0147086C65DE11030 ,pages-per-block-mlc = 258,capacity = 64000000000,pages-in-read-verify = 88,caus = 4,firmware-version = 12.22.01,sec-per-full-band-slc = 2752,ce-per-bus = 2,bytes-per-sec-meta = 16,num-dip = 8,dies-per-channel = 2,package_blocks_at_EOL = 16304,nand-marketing-name = 1Y128G-TLC-2p ,sec-per-full-band = 8256,cau-per-die = 2,page-size = 16384,pages-per-block-slc = 86,sec-per-page = 4,num-bus = 2,block-pairing-scheme = 0,chip-id = S3E,blocks-per-cau = 2108,Encryption Type = AES-XTS,vendor-name = Hynix ,pages-per-block0 = 0,default-bits-per-cell = 3,manufacturer-id = 0147086C65DE11030 }



http://forums.macrumors.com/threads/iphone-6s-64gb-tlc-nand-and-nvme-pcie-controller.1922221/

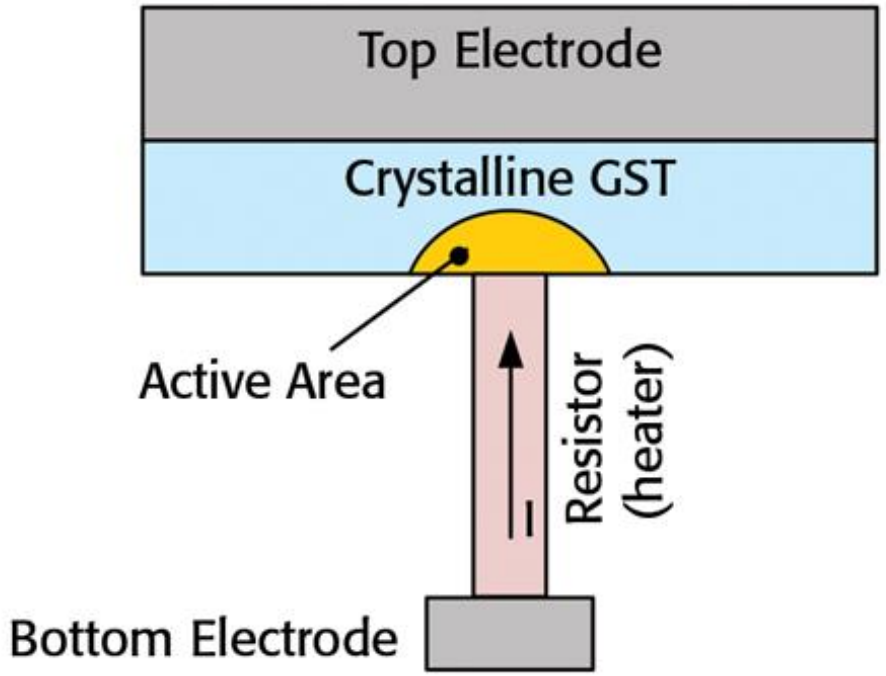
Разные xRAM



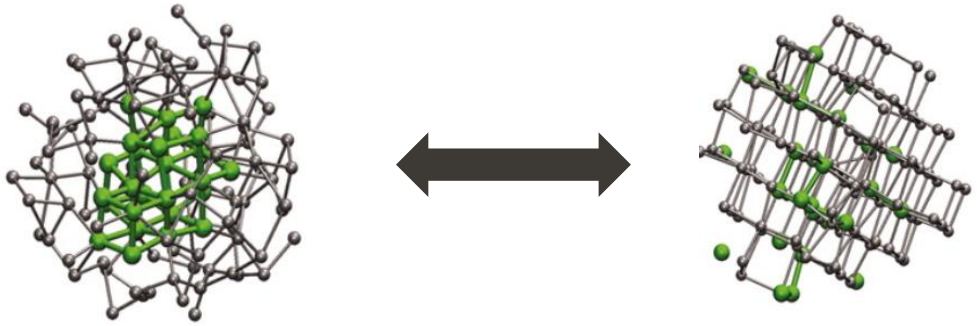
Dr Seshubabu Desu, 4DS Memory

Phase Change Memory (PCM, PRAM)

- Изменение агрегатного состояния под действием температуры

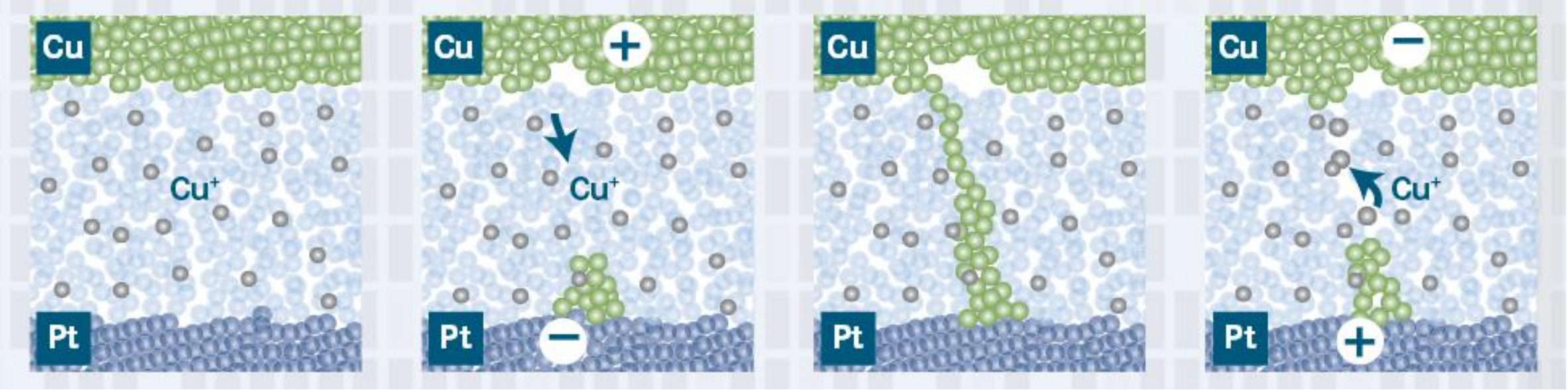


- Используется специальный класс веществ - халькогенидные стёкла ($Ge_2Sb_2Te_5$ (GST))
- Также используется в CD/DVD-RW за счет изменения отражающей способности
- Сопротивление в аморфном состоянии больше, чем в кристаллическом 1:100 - 1:1000

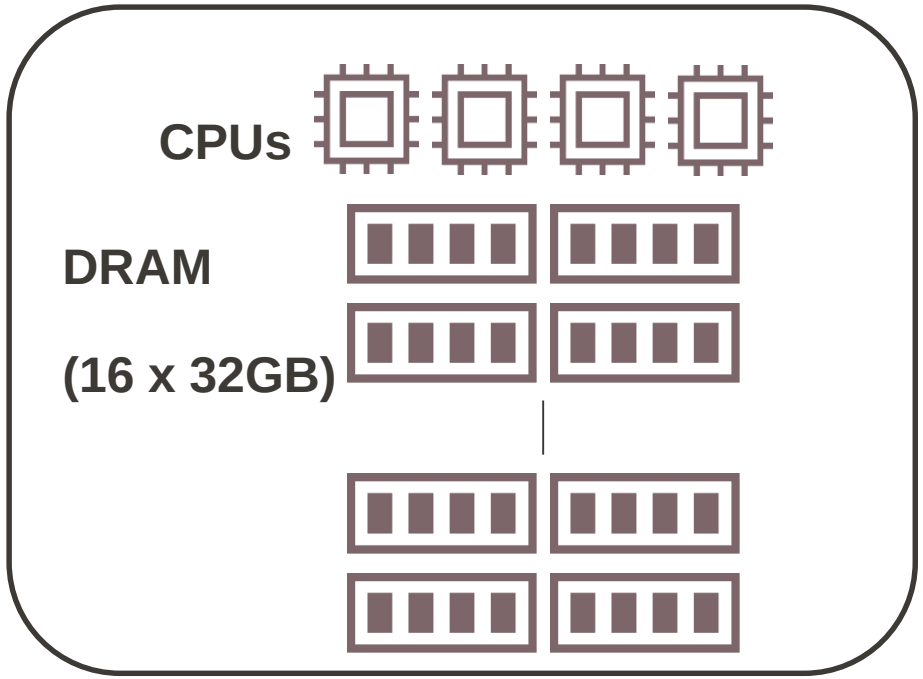


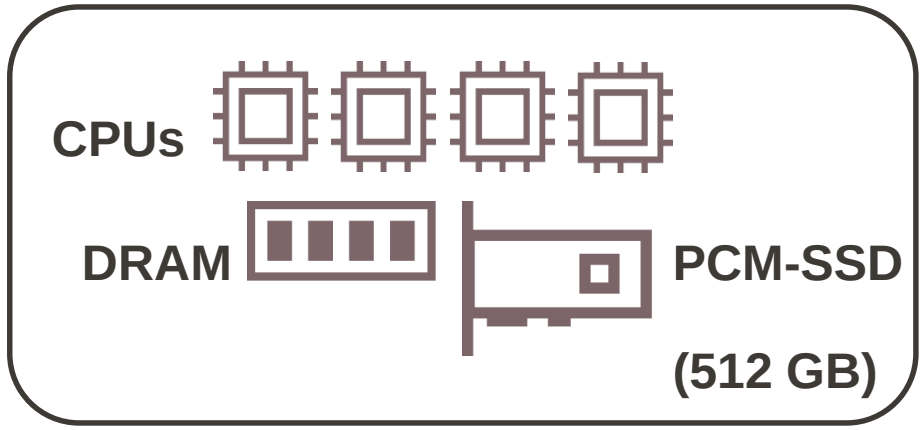
ReRAM / CBRAM

- Изменение характеристик диэлектрика (MeOx) за счет повышения напряжения.
- Формирование проводящей нити – ON/SET
- Разрушение проводящей нити – OFF/RESET



Сервер 2018-20





DRAM $8/GB
PCM $4/GB
3D NAND $0.2/GB

- 512GB memory = 16 x 32GB DIMM
- Питание на refresh памяти: 45W
- Максимальное потребление: 85 W/512GB

- 512GB PCIe card
- Питание на refresh : 0 W
- Максимальное потребление: 25W/PCIe slot

Серия Ultrastar® SN100

Твердотельные накопители NVM Express™

Беспрецедентное увеличение скорости работы приложений

- **Лидер по производительности**
  - 740K IOPs random read, 310K IOPs random r/w mix 70/30
  - Тесты SSD CERN: http://lvalsan.web.cern.ch/lvalsan/lfTnke7TrHowZmQxu1hsqI9Iuhg5LVr/
- **Различные форм-факторы**
  - HH-HL или 2.5” SFF
- **Емкость до 3.2ТБ**
  - 800ГБ, 1.6ТБ, 3.2ТБ для 2.5” SFF, 1.6ТБ и 3.2ТБ для PCIe HH-HL
- **Возможность переформатировать на больший объем или производительность**

HGST Ultrastar™ SN200 Series
NVMe PCIe SSD

2nd Gen

2nd Gen NVMe - 2x the capacity and 2x the AIC read performance (6.1GB/s vs 3.0GB/s) vs. SN100

7.68TB

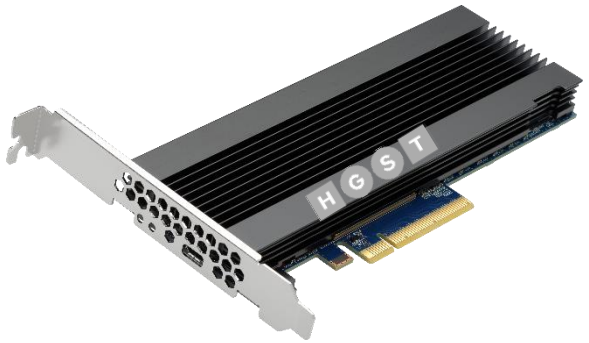
Industry's highest capacity NVMe SSD in 2.5" and HH-HL

1.2M

Industry's highest performance in 4K Random Reads up to 1.2 million IOPS

560K

Exceptional in mixed workloads (4K 70/30) up to 560K 4KiB IOPS



Ultrastar™ SN200 Series NVMe PCIe SSD

Skyhawk™ – Essential PCIe SSD

- Initial product into entry level NVMe space ('Essential' NVMe)
- Higher performance than SATA provides upgrade path to better access density
- 12W power profile eliminates cooling concerns of higher performance devices
- Leveraged platform: Same hardware and FTL as the Orion/Odyssey platform
- Built using SanDisk 15nm MLC NAND flash technology
- Better (lower) latency and optimized QoS compared to SATA SSDs

nvm EXPRESS

"12W SATA replacement PCIe SSD targeted delivering > 2X performance of SATA"

Form Factor & Power	JEDEC Endurance (DW/D for 5 years)	User Capacities (GB)
2.5" 15mm (12W)	1.7/1.2 DW/D	1600/3200
	0.6/0.5 DW/D	1920/3840

Notes: Skyhawk™ DW/D is for 5 years per JESD219 specifications;

Сравнение NVMe (RI)

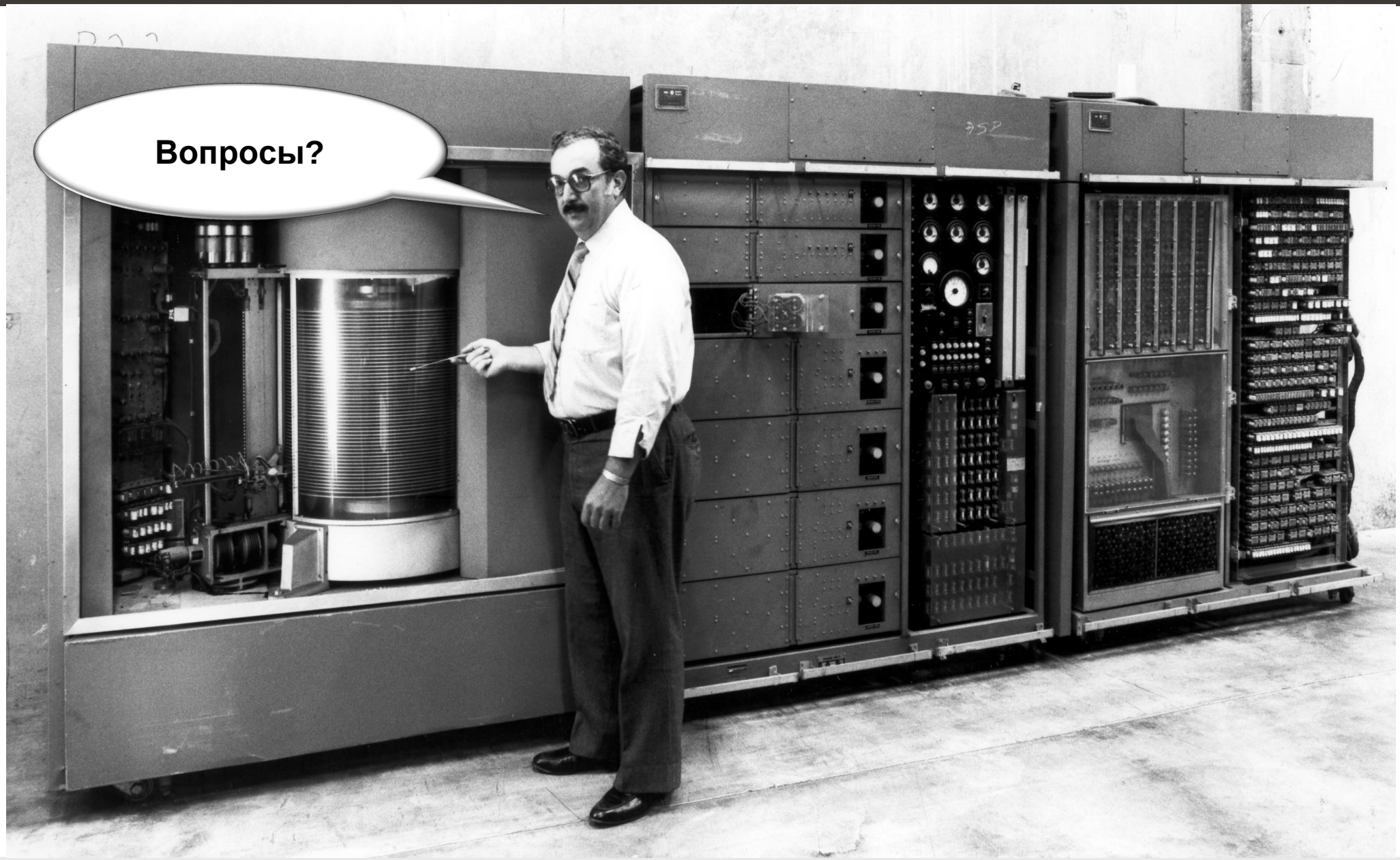
		SN100	SN200	Skyhawk Standard
Форм-фактор		AIC, U.2	AIC, U.2	U.2
Интерфейс		PCIe 3.0 x4	PCIe 3.0 x8, PCIe 3.0 x4	PCIe 3.0 x4
Объем		956GB[1], 1.91TB, 3.82TB	960GB[1], 1.92TB, 3.84TB, 7.68TB	1.92TB, 3.84TB
Технология NAND		A19 nm	15 nm	15 nm
DWPD (JESD219)		0.8 DWPD	1 DWPD	0.5 DWPD
Скорость	Seq Read/Write (GB/s)	3.0 / 1.6	6.1 / 2.2	1.5 / 1.17
	Rand Read/Write 4KB (IOPS)	743K / 38K	1.2M / 75K	250K / 47K
	Rand Read/Write/70-30 4KB (IOPS)	115K	270K	99K
Потребление		25W	25W	10W
Надежность (UBER)		1 на 10^{17}	1 на 10^{18}	1 на 10^{17}

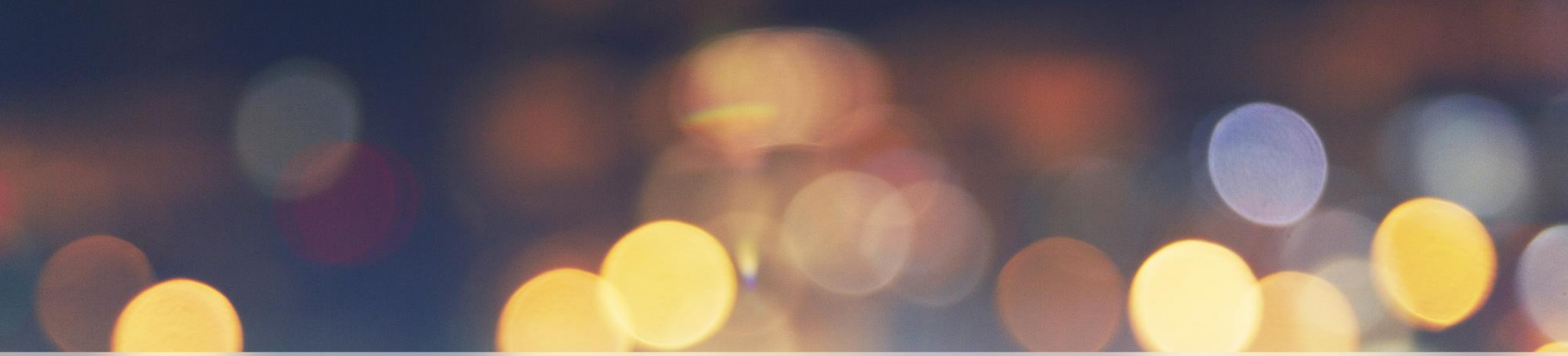
1: только для U.2 SFF

Сравнение NVMe (MU)

		SN100	SN200	Skyhawk Ultra
Форм-фактор		AIC, U.2	AIC, U.2	U.2
Интерфейс		PCIe 3.0 x4	PCIe 3.0 x8, PCIe 3.0 x4	PCIe 3.0 x4
Объем		800GB[1], 1.6TB, 3.2TB	800GB[1], 1.6TB, 3.2TB, 6.4TB	1.6TB, 3.2TB
Технология NAND		A19 nm	15 nm	15 nm
DWPD (JESD219)		3 DWPD	3 DWPD	1.2/1.7 DWPD
Скорость	Seq Read/Write (GB/s)	3.0 / 1.6	6.1 / 2.2	1.7 / 1.2
	Rand Read/Write 4KB (IOPS)	743K / 140K	1.2M / 200K	250K / 83K
	Rand Read/Write/70-30 4KB (IOPS)	310K	560K	150K
Потребление		25W	25W	10W
Надежность (UBER)		1 на 10^{17}	1 на 10^{18}	1 на 10^{17}

1: только для U.2 SFF


Вопросы?



Спасибо!