Архитектура вычислительных систем

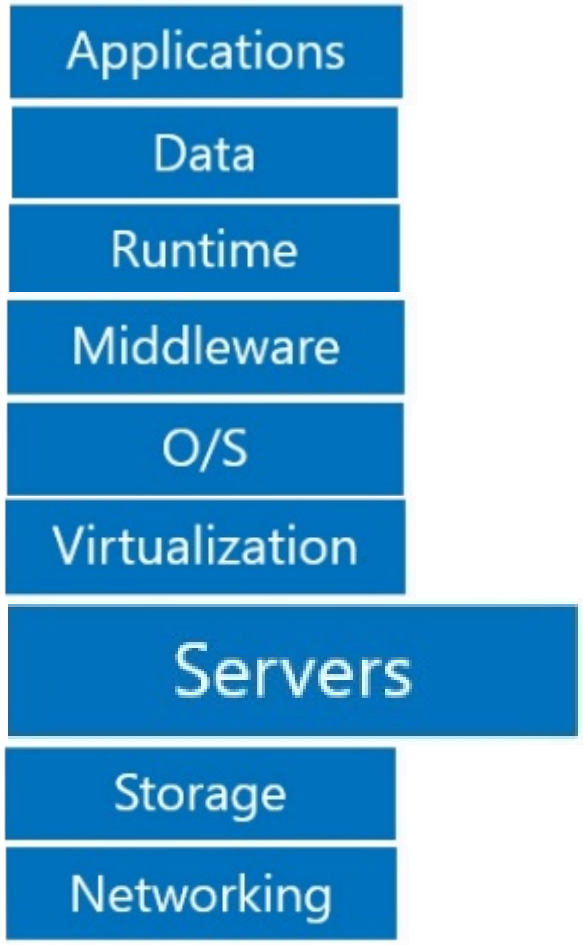
Лекция 2. Аппаратное обеспечение. Часть 1?

Artem Beresnev

t.me/ITSMDao

t.me/ITSMDaoChat

ИТ-инфраструктура



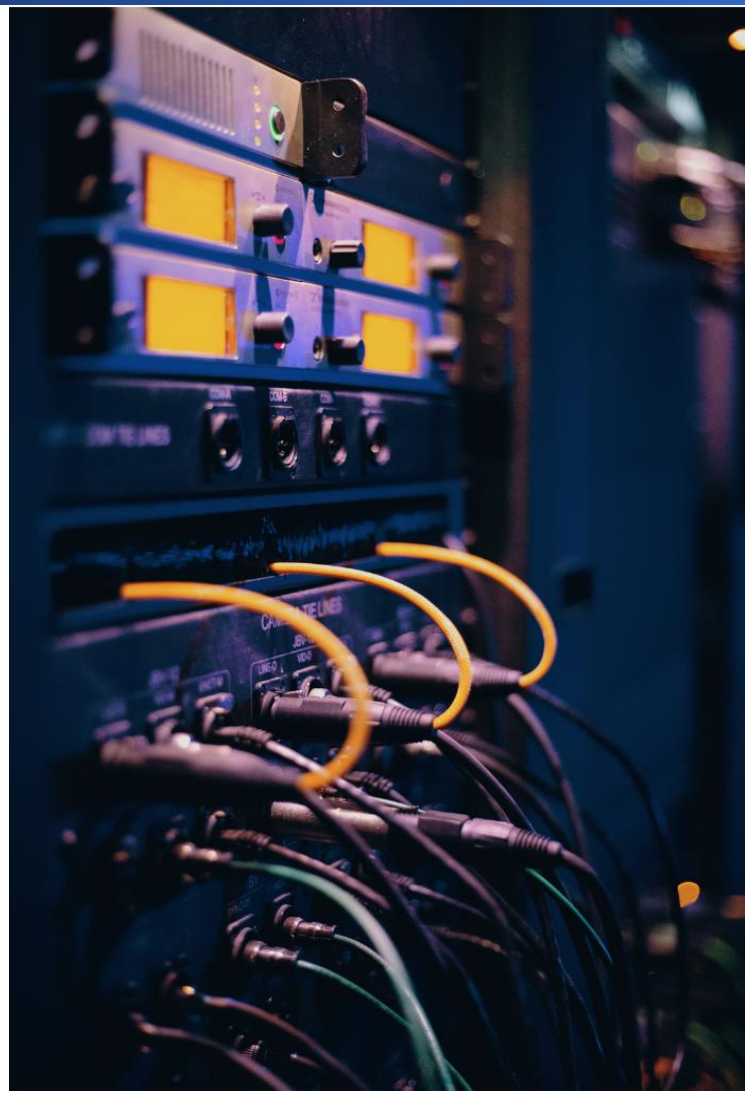
Это аппаратные платформы для вычислений, запуска программных компонентов.

План

- История архитектур
- Модульная архитектура и SoC
- CPU
- GPU
- Материнские платы
- Память
- Форм-факторы серверных платформ

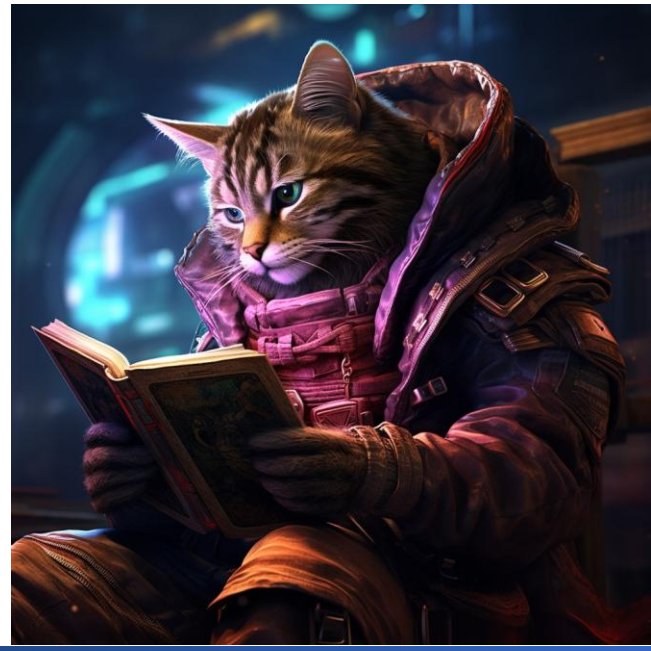
Архитектура вычислительных систем

Прежде чем переходить к конкретике, нужно обсудить основные идеи, лежащие в основе существующих платформ



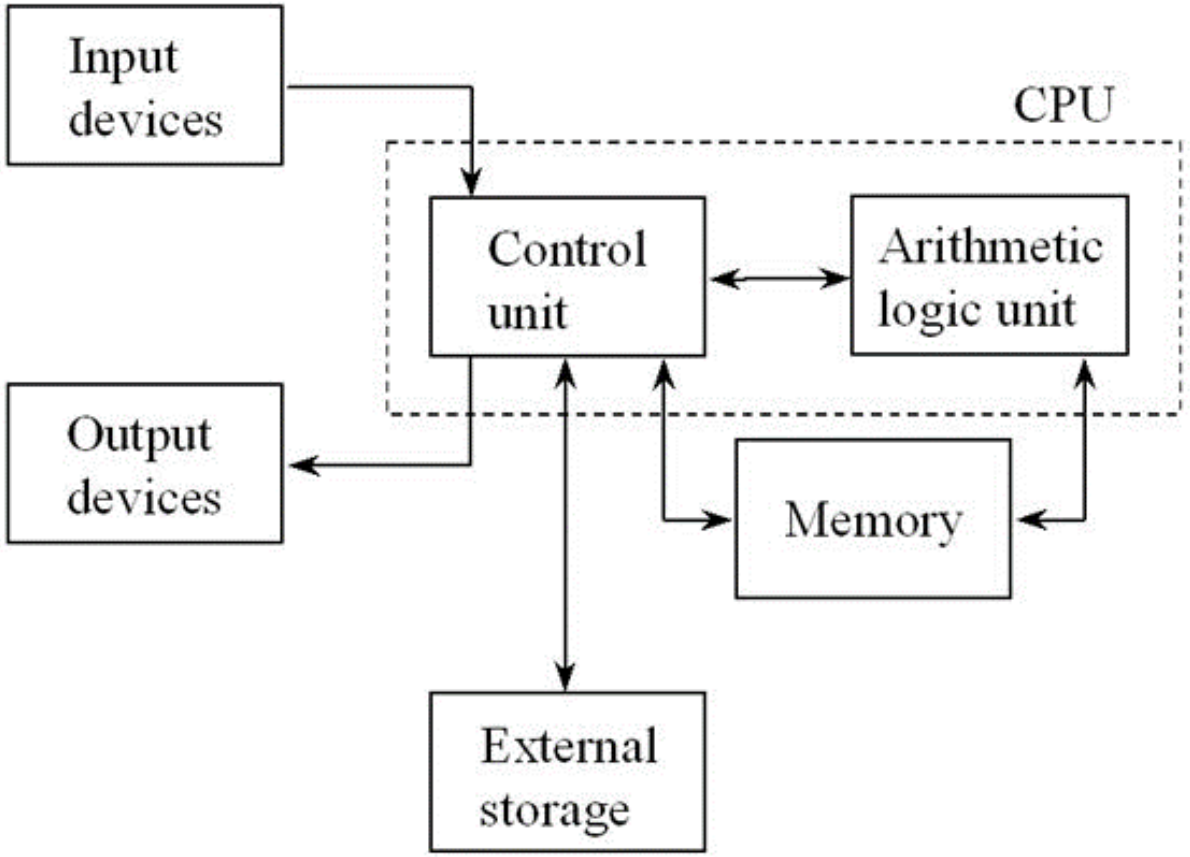
Немного истории

- Джон фон Нейман (1945)
  - Описана архитектура современных программируемых компьютеров.
  - Показано, что программы можно изменять, не меняя аппаратной части.
- Изобретение транзистора (1947)
- Изобретение БИС (средина 60-х)
- Изобретение микропроцессора (начало 70-х)



- Принцип открытой архитектуры (IBM)

Состав платформы



- Вычислитель
- ОЗУ
- ПЗУ
- Ввод\вывод
- Шины передачи данных

Подходы к хранению в памяти

По принципу использования памяти (или по структуре архитекторы):

- **Принстонская архитектура** (фон Неймана) – общая шина данных для обращения к памяти, где храниться и данные и команды. Плюсы – эффективное использование памяти, возможности манипуляции с командами. Минус- ограничения производительности по шине данных.

- **Гарвардская архитектура** - память программ и памяти данных физически разделена. Плюсы – быстродействие. Минусы- несколько шин, ограничения по размеру памяти.

Подходы к хранению в памяти

- **Принстонская архитектура** – используется во внешней структуре большинства процессоров и, следовательно, платформ.

- **Гарвардская архитектура** - во внутренней структуре современных высокопроизводительных микропроцессоров, где используется отдельная кэш-память для хранения команд и данных.

Компоновка компонентов

Модульные системы (IBM PC)

1981

из чего все состоит?

Модульные системы (IBM PC)

- CPU (*N)
  - Материнская плата
- Память
  - Диск(*2)
- Блок питания (*2)
  - Видеокарта(*2)
- Корпус
  - Сетевые интерфейсы
- Контроллеры дисковой подсистемы
  - Корпус

Модульные системы (IBM PC)

В центре – Материнская плата

Шины, ПЗУ, цепи питания, тактовые генераторы и прочия, прочия.

ДАВНО существовали
Северный мост (Northbridge) – отдельный чип для взаимодействия с оперативной памятью и графическими интерфейсами (AGP или PCI Express) и **Южный мост** (Southbridge) - чип отвечал за взаимодействие с медленными периферийными устройствами и интерфейсами, такими как SATA, USB, аудио, сети и прочие встроенные контроллеры.

Модульные системы (IBM PC)

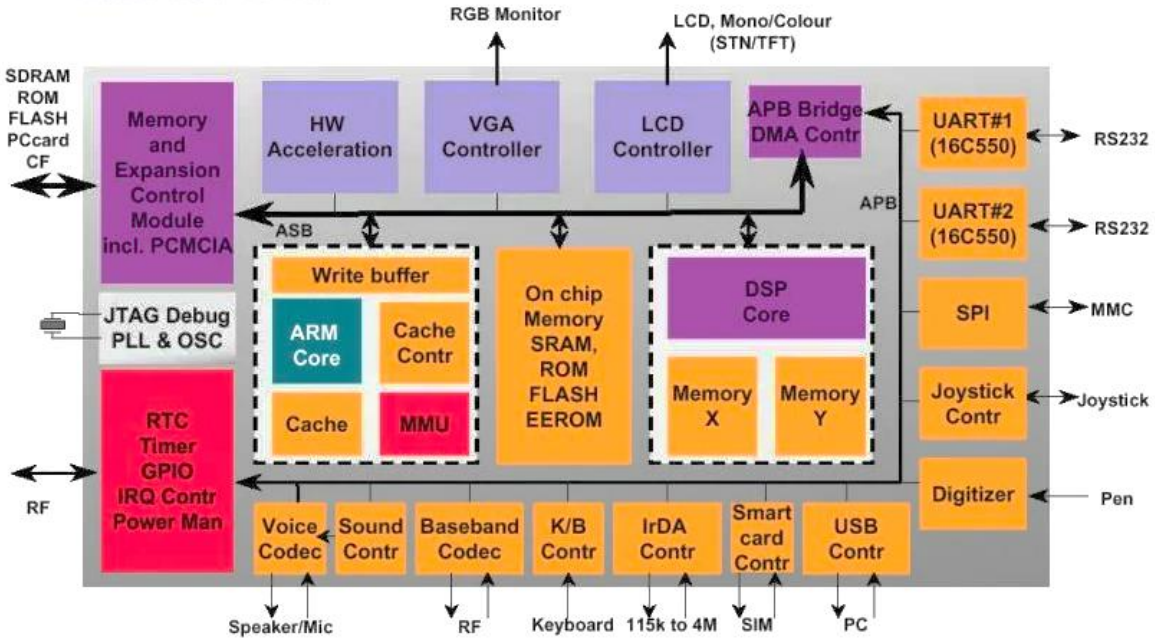
В центре – Материнская плата

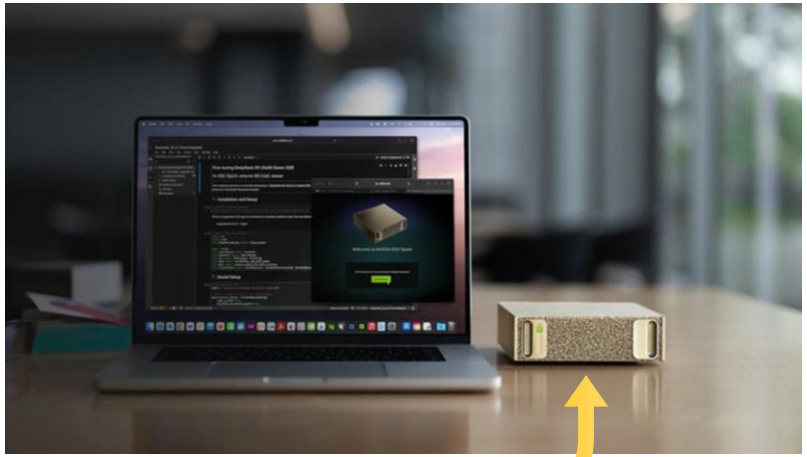
Шины, ПЗУ, цепи питания, тактовые генераторы и прочия, прочия.

СЕЙЧАС
северный мост в основном интегрирован в сам процессор
(Intel с Nehalem и для AMD с Fusion),
южный мост есть и сейчас как:
Platform Controller Hub (PCH) у Intel
FCH (Fusion Controller Hub) у AMD.

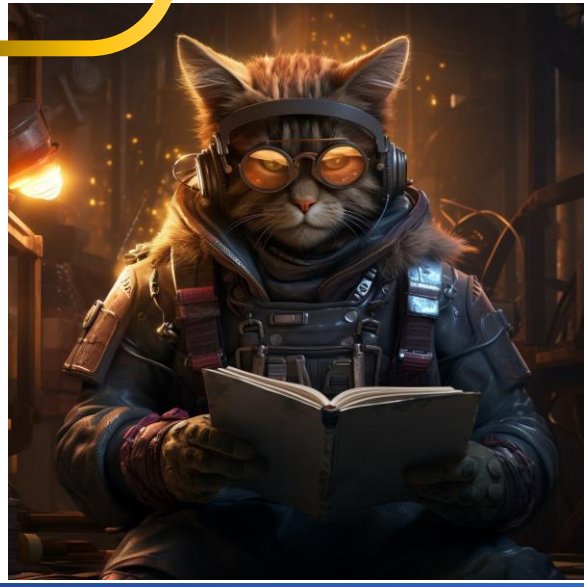
Компоновка компонентов

System On a Chip (SoC)



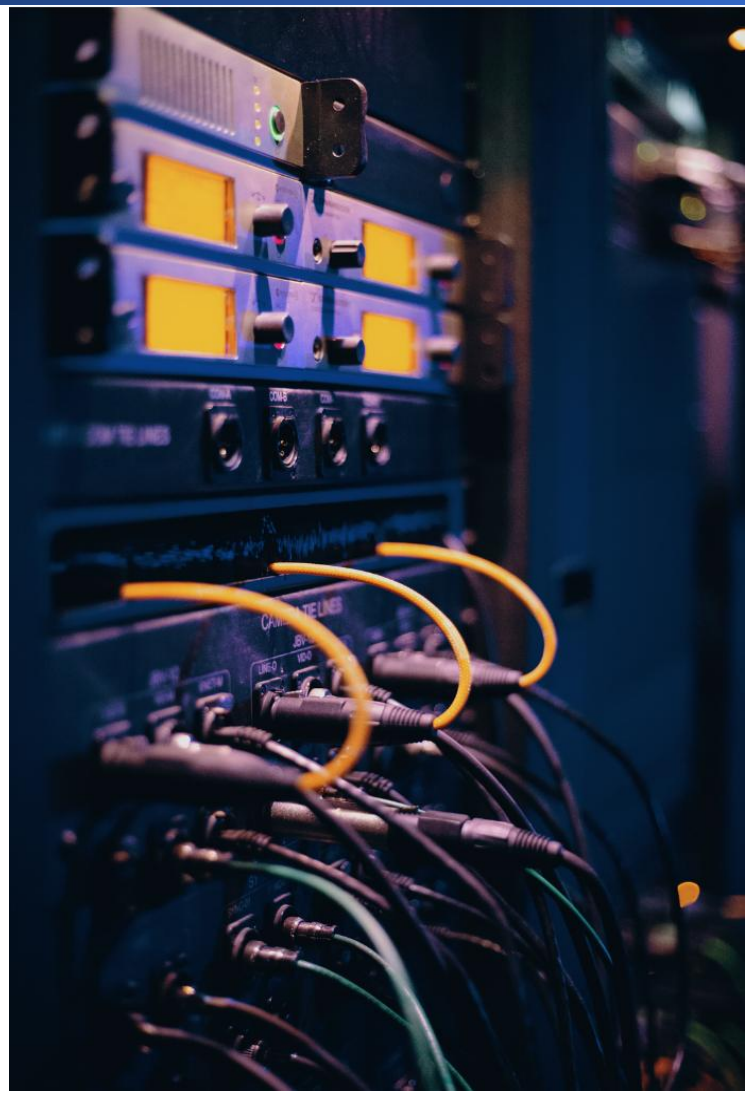


NVIDIA Spark



Подходы к внутренней архитектуре процессора

Процессоры - сложные устройства и их задачи можно решать по разному. Рассмотрим основные подходы.



Архитектура процессора

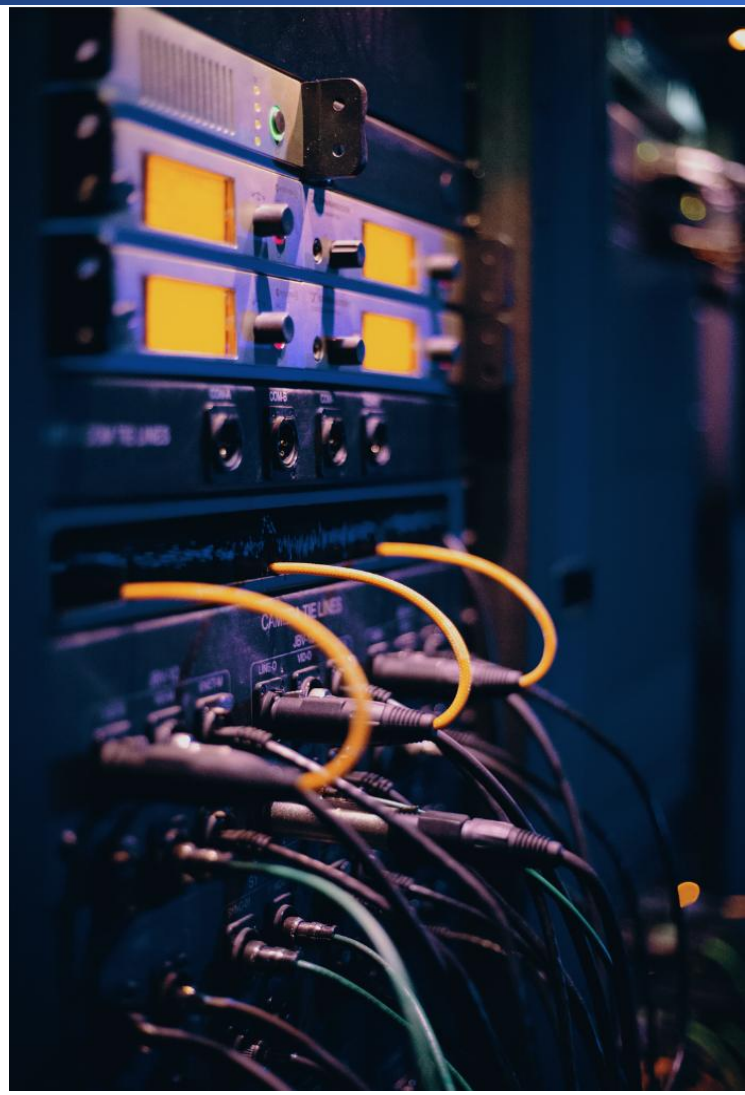
- **Архитектура как подход к разработке процессора**
  CISC, RISK, MISK, VLIW
- **Архитектура как набор команд (ассемблер)**
  Intel 64, AMD64, EM64T, x86-64, ARM9, ARM64.
- **Архитектура как это набор свойств и качеств, присущий одному семейству процессоров**

Архитектура процессора

- **CISC** (англ. Complex Instruction Set Computer — компьютер с полным набором команд
- **RISC** (англ. Reduced Instruction Set Computer — компьютер с сокращённым набором команд)
- **MISC** (англ. Minimal Instruction Set Computer — компьютер с минимальным набором команд)
- **VLIW** (англ. Very Long Instruction Word — очень длинная машинная команда)

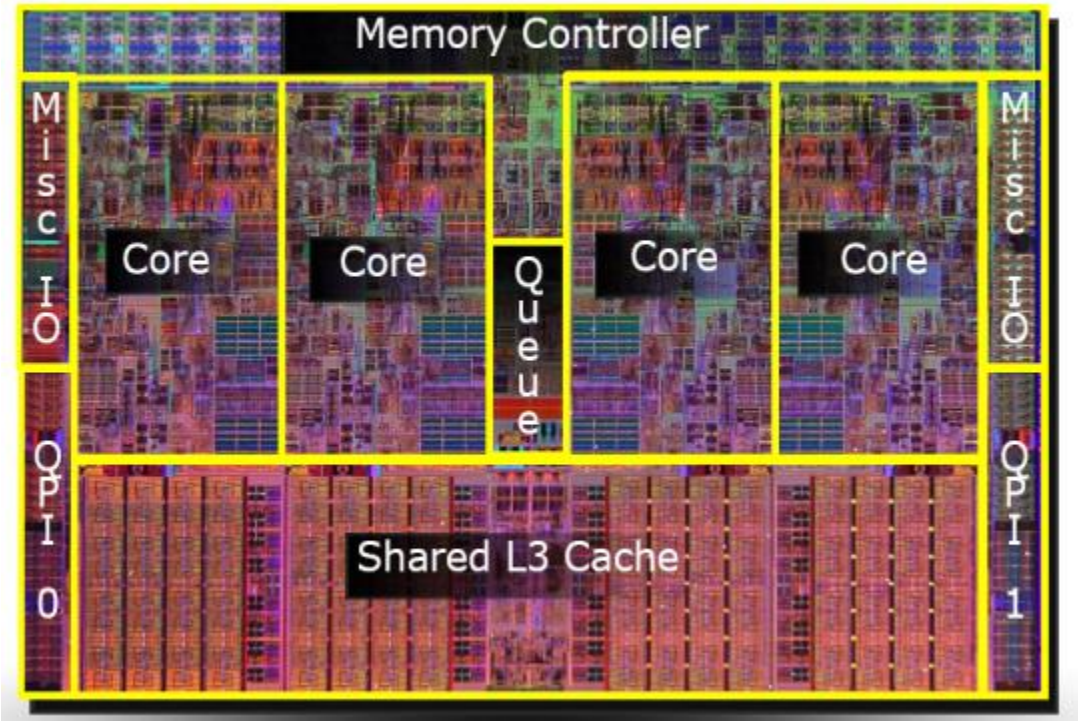
Немного о реальных CPU

Приведем примеры реальных CPU разных архитектур и опишем некоторые их компоненты и технологии



Топология

- Количество:
  - процессоров
  - физических ядер
  - гиперпотоков в ядре
- Их произведение – количество логических ядер
- APIC (advanced programmable interrupt controller) — это устройство входящее в состав процессора, отвечающее за работу с прерываниями, приходящими к конкретному логическому процессору. Свой собственный APIC ID есть у каждого логического процессора.
- Проблема управления со стороны ОС

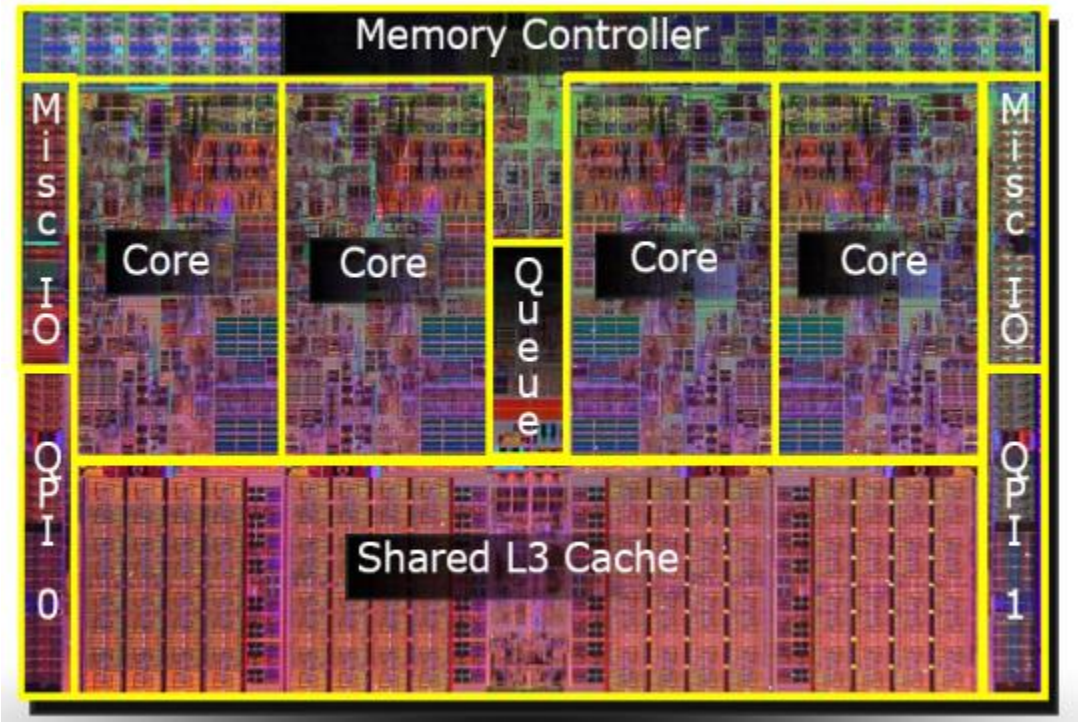


Топология

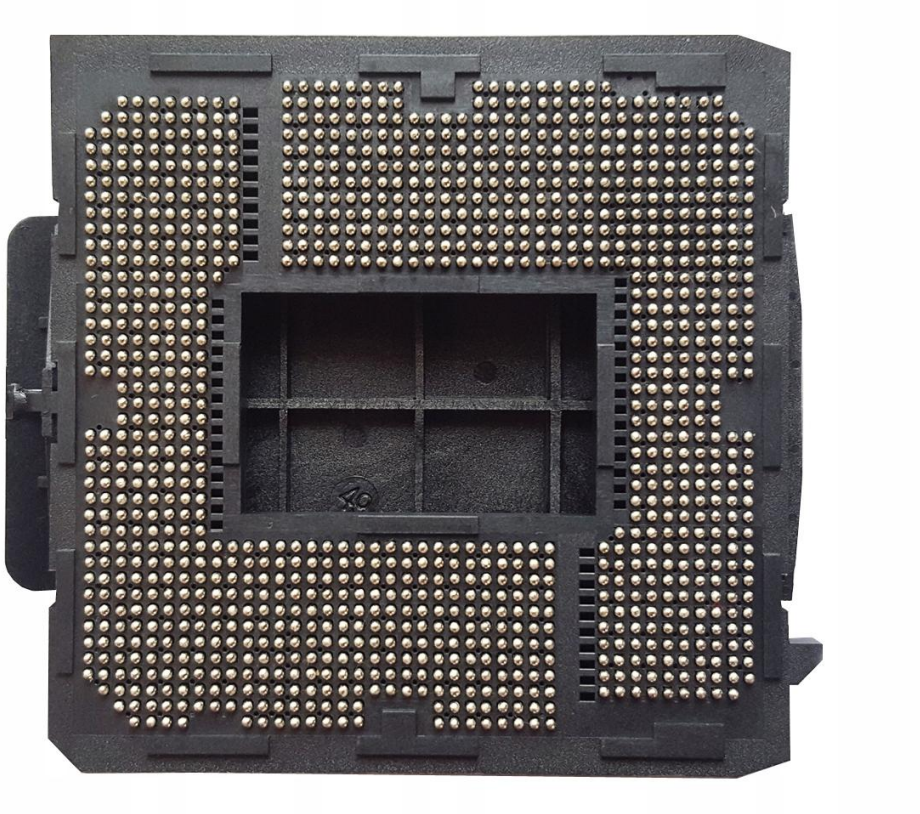
/proc/cpuinfo

- processor: ID указывающее логический процессор (включая ядра и потоки), начиная с нуля.
- physical id: ID физического процессора (или сокета) на материнской плате. Если у системы несколько ЦП (например, в системе с несколькими процессорными сокетами), у каждого физического процессора будет свой physical id.
- siblings: количество логических процессоров (потоков), которые доступны на данном физическом процессоре.
- core id: идентификатор ядра на физическом процессоре.
- cpu cores: количество физических ядер на процессоре.
- apicid: идентификатор в Local APIC (Advanced Programmable Interrupt Controller)
- initial apicid: начальный APIC ID до любого изменения со стороны BIOS/OS.

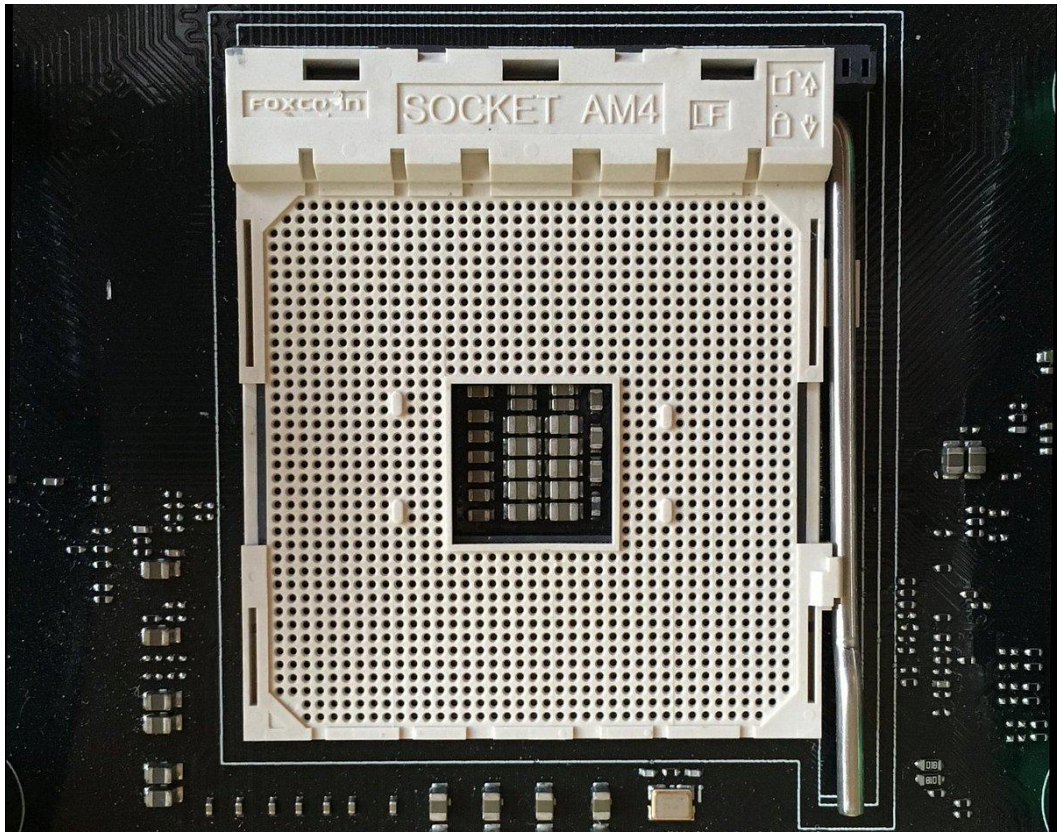
taskset



Сокеты



LGA 1151



Socket AM4

Кэш

- Кэш L1 (инструкции) - Характеристика указывает объем кэш-памяти первого уровня, данного процессора. L1 делится на кэш данных (L1D) и кэш команд или инструкций (L1I). Принадлежит только конкретному ядру процессора. Типичные размеры: от 8 до 384 КБ.
- Кэш L2 - Характеристика указывает объем кэш-памяти второго уровня, данного процессора.
- Кэш L3 - Кэш третьего уровня наименее быстродействующий, но он может быть очень большим — более 24 Мбайт. L3 медленнее предыдущих кэшей, но всё равно значительно быстрее, чем оперативная память.

Частота и TDP

- Частота

- Свободный множитель процессора позволяет изменять его тактовую частоту стандартными средствами материнской платы и чипсета. Наличие свободного множителя необходимо для разгона процессора.

- Тепловая мощность проектирования (**TDP**) - максимальное количество тепла, генерируемое компьютерным чипом или компонентом (часто процессором, графическим процессором или системой на кристалле), которое система охлаждения компьютера предназначена рассеивать при любой нагрузке.

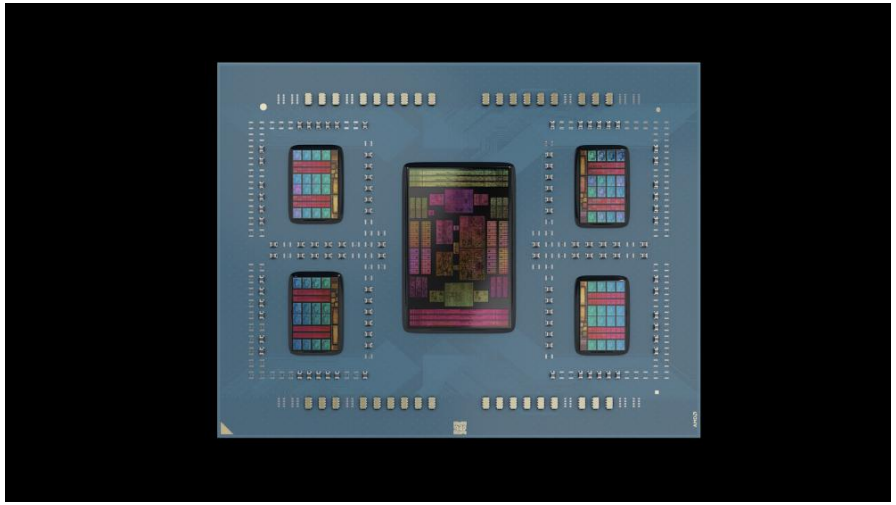
Конвейер и предсказание ветвления

- Конвейер команд — это метод, используемый в архитектуре CPU (центрального процессора) для увеличения производительности, позволяя нескольким инструкциям находиться на разных стадиях выполнения одновременно (выборка инструкции, декодирование, выполнение, доступ к памяти и запись результата).

- Аппаратное предсказание ветвлений — сложная система, предназначенная для улучшения эффективности конвейерной обработки команд, минимизируя задержки, вызванные ветвлениями в коде (предвыборка, упорядочивание, спекулятивное выполнение)

Примеры AMD EPYC

Model	Fab	Cores (Threads)	Chiplets	Core config[i]	Clock rate (GHz)		Cache (MB)			Socket	Socket count	PCIe 5.0 lanes	Memory support	TDP	Release date	Price (USD)
					Base	Boost	L1	L2	L3				DDR5 ECC			
Cloud (Zen 4c cores)																
9734	TSMC N5	112 (224)	8 × CCD 1 × I/OD	8 x 14	2.2	3.0	7	112	256	SP5	1P/2P	128	DDR5-4800 twelve-channel	340 W	Jun 13, 2023	$9,600
9754S		128 (128)		8 x 16	2.25	3.1	8	128						360 W		$10,200
9754		128 (256)														$11,900



Примеры Intel Xeon

Model number	Cores (threads)	Base clock	Turbo Boost		Smart cache	TDP	Maxi-mum scala-bility	Registered DDR5 w. ECC support	UPI links	Release MSRP (USD)
			All core	Single core						
Xeon Platinum (8400)										
8490H	60 (120)	1.9 GHz	2.9 GHz	3.5 GHz	112.5 MB	350 W	8S		4	$17000
8488C	48 (96)	2.4 GHz	3.2 GHz	3.8 GHz	105.0 MB	385 W	2S		?	
8487C	56 (112)	1.9 GHz	?	3.8 GHz		350 W			?	
8481C		2.0 GHz	2.9 GHz						?	
8480+			3.0 GHz						4	$10710
8480C										
8478C	48 (96)	2.2 GHz	?						?	
8475B		2.7 GHz	3.2 GHz		97.5 MB				?	
8474C		2.1 GHz	?						?	
8473C	52 (104)		2.9 GHz		105.0 MB				?	
8471N		1.8 GHz	2.8 GHz	3.6 GHz	97.5 MB	300 W	1S		4	$5171
8470Q		2.1 GHz	3.2 GHz	3.8 GHz	105.0 MB	350 W	2S			$9410
8470N		1.7 GHz	2.7 GHz	3.6 GHz	97.5 MB	300 W				$9520
8470		2.0 GHz	3.0 GHz		105.0 MB	350 W				$9359

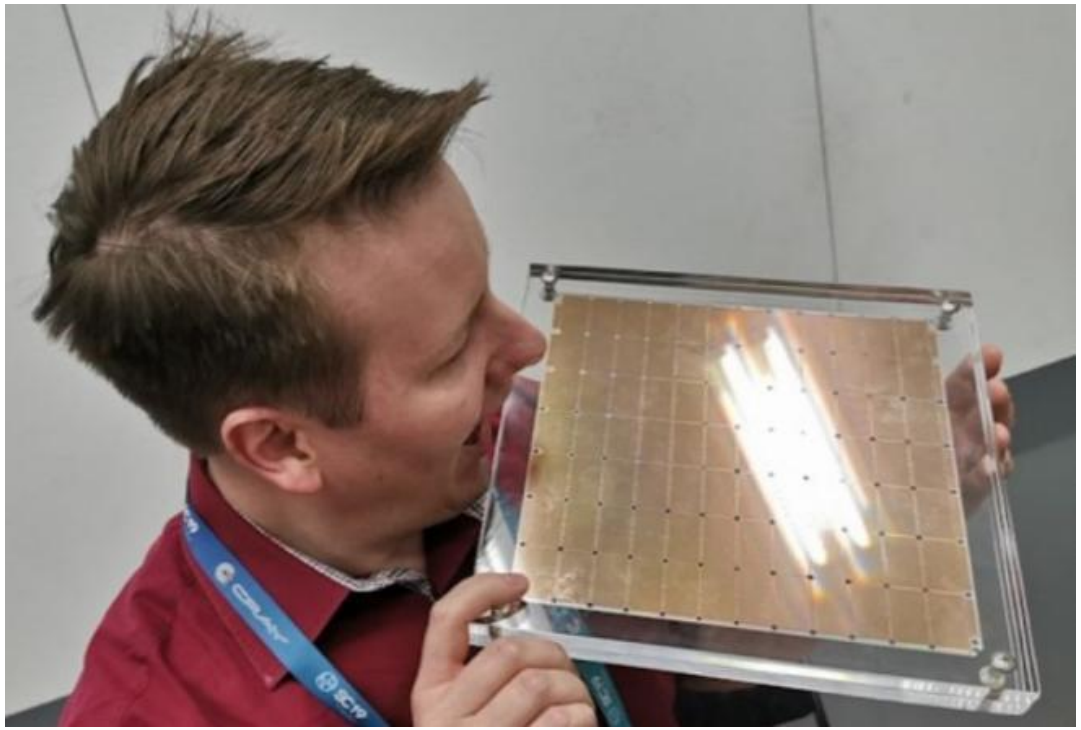
Примеры Apple

Примеры Apple

Характеристики	Apple M1	Apple M2
Техпроцесс	5 нм	5 нм (Gen2. N5P)
Количество транзисторов	16 млрд	20 млрд
Площадь кристалла	120,5 мм. кв	142 мм.кв.
Производительные ядра (Firestorm)	4	4
Энергоэффективные (Icestorm)	4	4
Кэш L2 (произв. ядра + энергоэффективность ядра)	12+4 Мб	16+4 Мб
Кэш L3	16 мб	8 мб
Максимальная базовая частота ядер (произв./энергоэф.)	3.2 / 2.1 GHz	3.5 / 2.4 GHz
Система Neural Engine	16 ядер, 11 TOPS	16 ядер, 15.8 TOPS
Число ядер графического процессора, и его пропускная способность	7-8 ядер, 2.6 TFlops	10 ядер, 3.6 TFlops
Максимальный объем и тип поддерживаемой памяти	8-16Gb 128-bit LPDDR4-4266	8-16-24Gb 128-bit LPDDR5 6400
Пропускная способность памяти	68 Gb/s	100 Gb/s
Модули кодирования и декодирования	H. 264, H.265 4K	H. 264, H.265, ProRes RAW 8K
PCIe версия	4.0	4.0
USB версия	USB 4/Thunderbolt3 x2	USB 4/Thunderbolt3 x2
Дата выхода	Ноябрь 2020	Июнь 2022

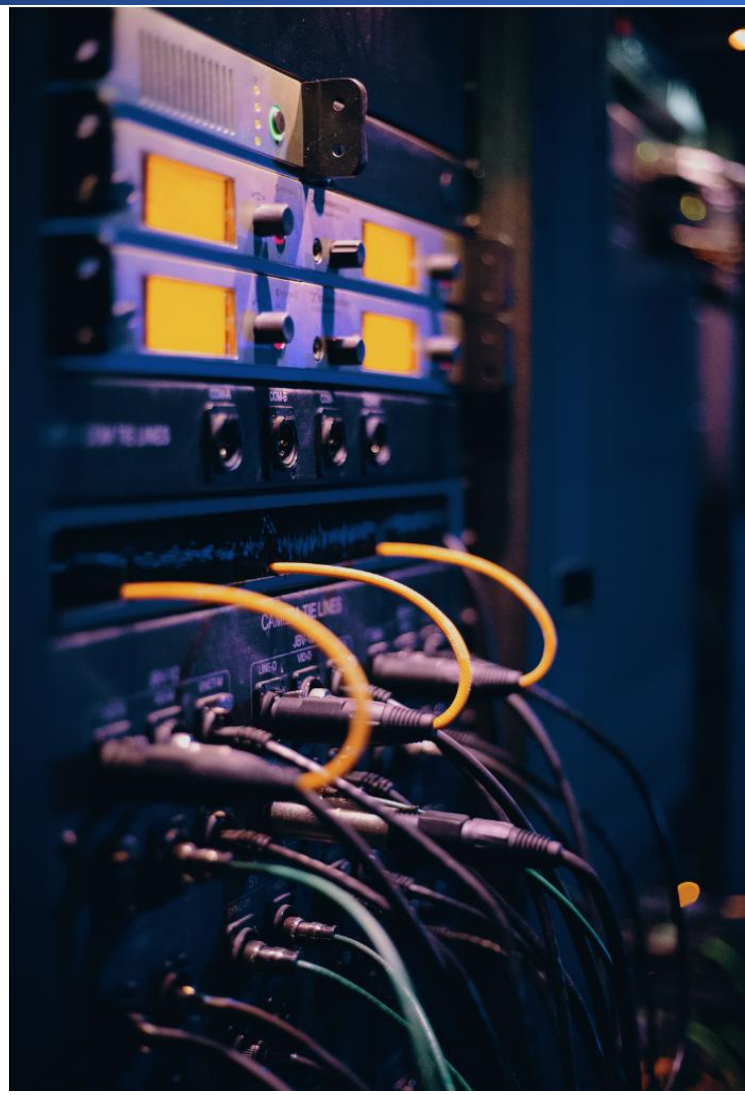
Что то действительно впечатляющее

- Процессор Wafer Scale Engine 3 от компании Cerebras
- На плате:
  - 900 000 ядер,
  - 44 ГБ памяти SRAM,
  - 4 триллиона транзисторов
- По мощности один процессор сравним с тысячей GPU NVIDIA V100
- Энергопотребление составляет 15 кВт.
- https://cerebras.ai/product-chip/



Как сравнивать процессоры?

Слишком они разные…

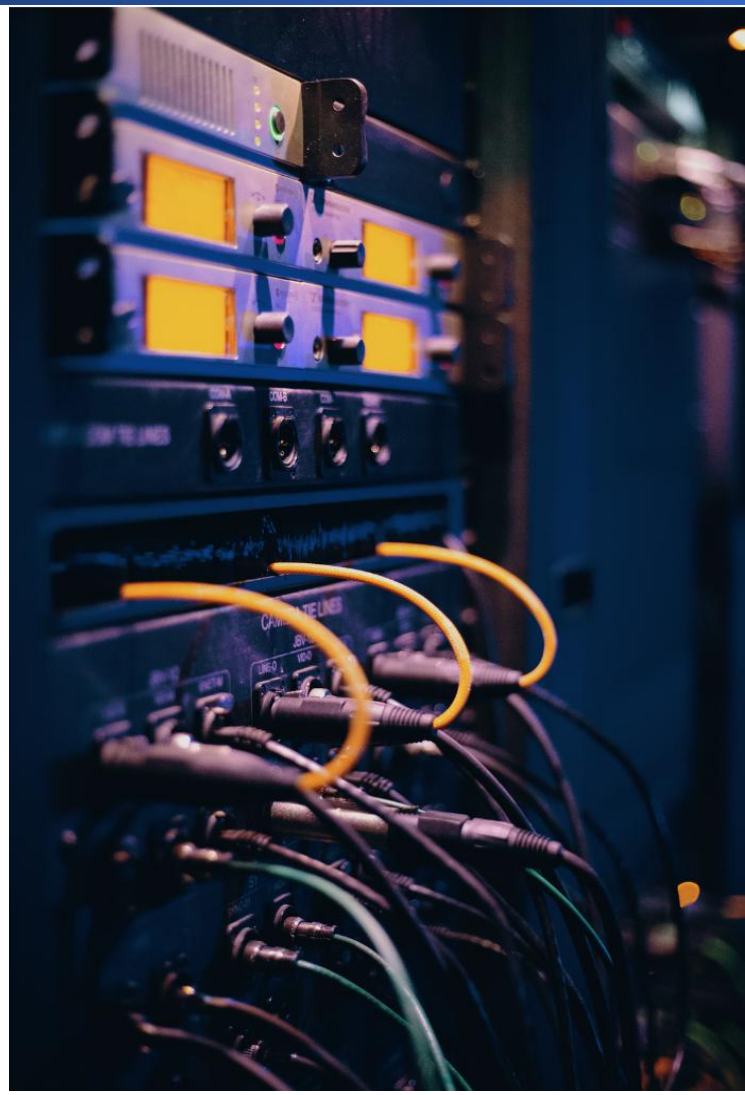


Сравнение CPU

- Класс
- Формальные показатели
- Benchmark (https://versus.com/ru/cpu)
- Цена \ TCO (Total Cost of Ownership) \ ROI (Return on investment)

GPU

Современные вычислительные платформы немыслимы без GPU. Опишем их особенности и области применения



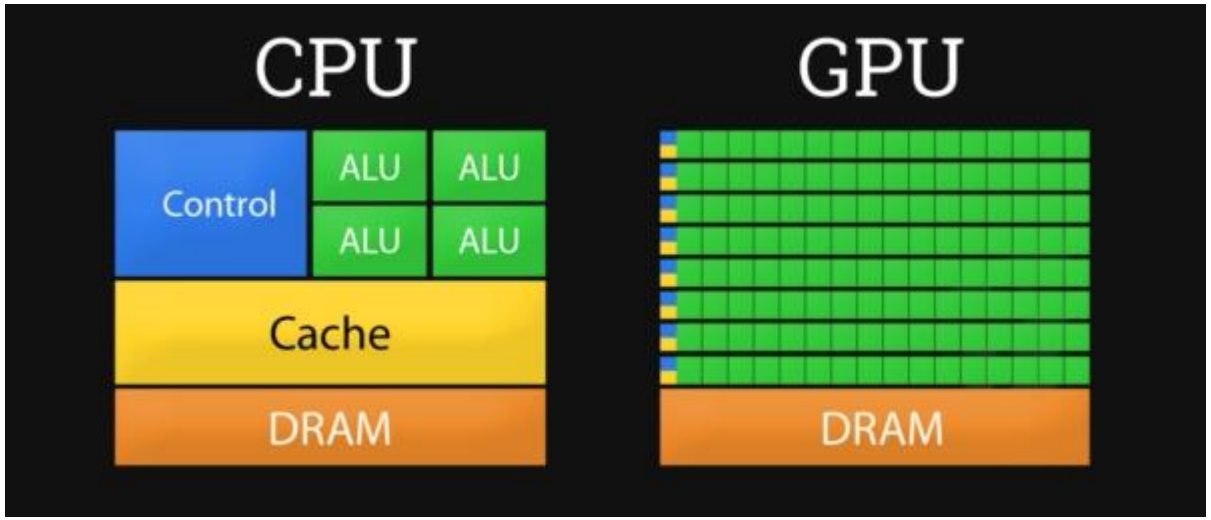
Отличия GPU от CPU

- Графические процессоры (GPU) имеют свою собственную архитектуру
- Архитектура GPU компании Nvidia называется CUDA (Compute Unified Device Architecture).
- Цель – обеспечить параллельную обработку данных и выполнения большого количества потоков одновременно, что идеально подходит для рендеринга графики, машинного обучения, симуляций и т.п.

Состав GPU

- Потоковые процессоры (ядра).
  Ядра работают параллельно, что обеспечивает высокий уровень параллелизма в вычислениях. Есть специализация ядер.
- Память
  Видеопамять (VRAM), которая предназначена для хранения больших объемов графических данных. VRAM включает несколько уровней кэша, таких как L1 и L2.
- Широкополосная шина данных
  GPU обладает широкой шиной данных, что позволяет быстро передавать большие объемы информации между потоковыми процессорами и видеопамятью.

Отличия GPU от CPU



- Сетка (GRID)
  - Блоки (thread blocks)
    - Ядра
    - Контрольное устройстов
    - Кэш (L1 для блока, shared mem, L2 – для GPU)

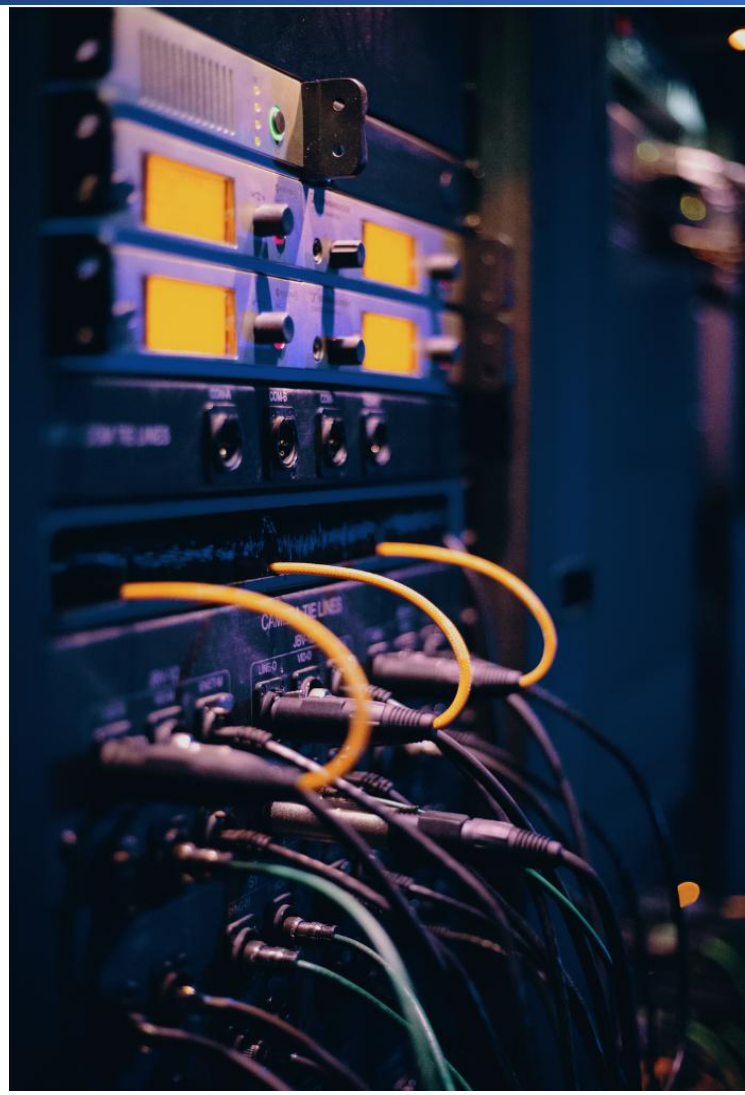
GPU

https://resources.nvidia.com/en-us-dgx-systems/dgx-h200-datasheet

https://www.nvidia.com/content/dam/en-zz/Solutions/design-visualization/quadro-product-literature/quadro-rtx-8000-us-nvidia-946977-r1-web.pdf

Материнские платы

Сам по себе процессор ничего не может. Его нужно обеспечить питанием и доступом к периферии



Материнские платы

- Поддержка сокета
- Питание и тактовые генераторы
- Поддержка памяти
- Интерфейсы и шины
- Микропрограммы

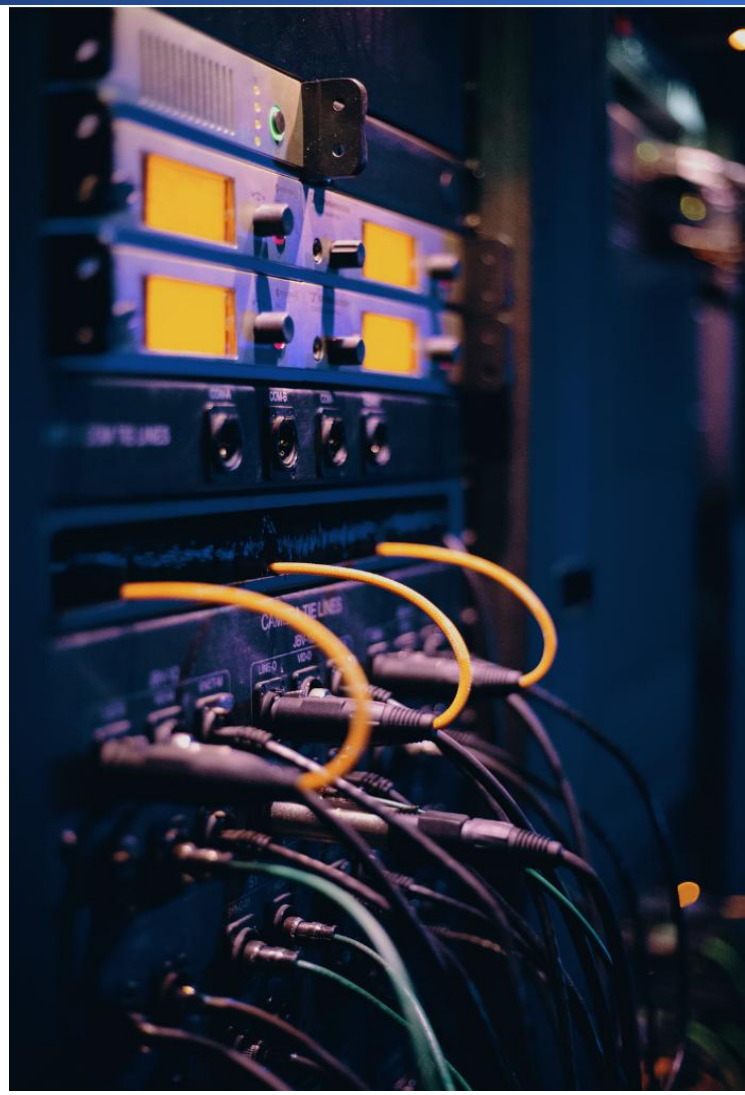
Материнские платы

Пример: MSI Z490-A PRO

- Сокет LGA 1200
- **Чипсет Intel Z490**
- DDR4 без ECC до 128 ГБ, частота от 2133 МГц до 4800 МГц
- 6x SATA 6Gb/s
- M2
- NVMe
- Количество слотов PCI-E x16 (CrossFire X)
- Количество слотов PCI-E x1
- Звук Realtek ALC892 \ Сеть - Realtek RTL8125-CG

Память

Важная часть системы, во многом определяющая быстродействие



Оперативная память

- Класс (desktop \ server)
- Тип памяти (DDR3, DDR4, DDR5)
- Объем, Частота, Тайминги.
- Форм-фактор (DIMM, SO-DIMM)
- Двухканальная компоновка

Тайминги оперативной памяти

- Тайминги показывают время (в тактах), которое проходит от момента отправки памятью команды и её фактическим исполнением (обычно 4 числа).
  - CAS Latency (CL - самый важный показатель) обозначает число тактов, которое проходит между отправкой запроса и началом ответа;
  - RAS to CAS Delay - число тактов, которое у контроллера занимает активация нужной строки банка;
  - RAS Precharge - число тактов, которое требуется для закрытия одной строки данных и перехода к другой;
  - Row Activate Time - число тактов до закрытия строки.

Тайминги оперативной памяти

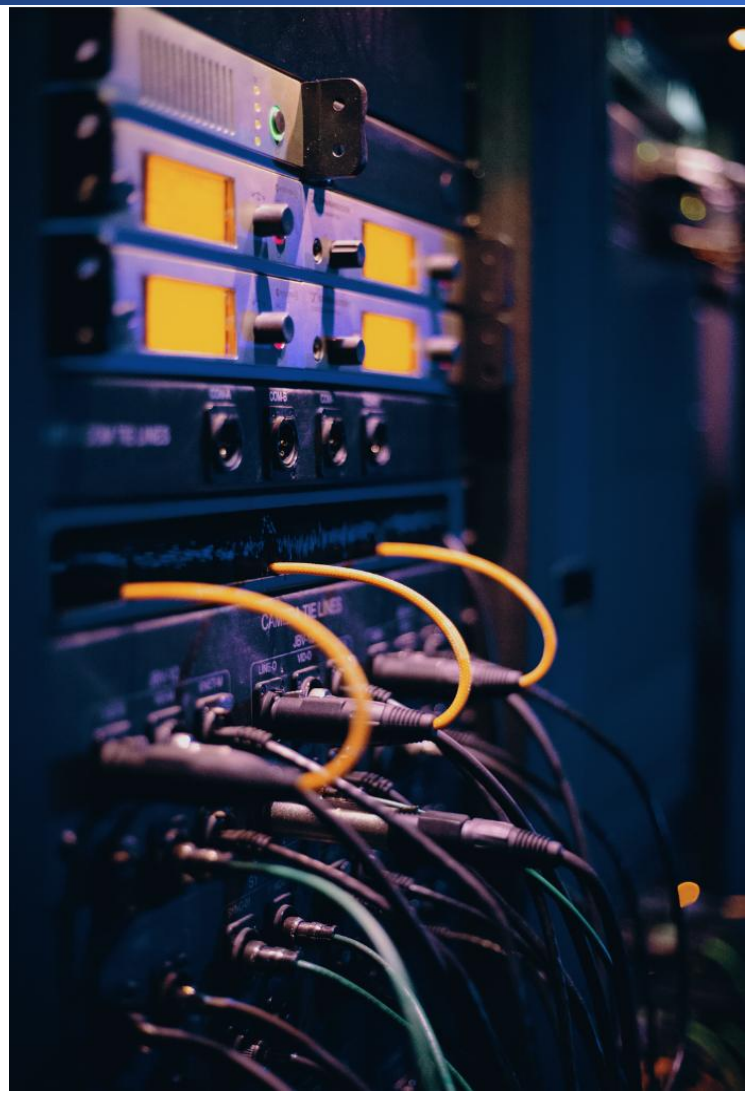
Характеристика	DDR3	DDR4	DDR5
Частота (МГц)	800–1600	2133–3200+	4800–8400+
Пропускная способность	~12–15 ГБ/с	~17–25 ГБ/с	~38–67+ ГБ/с
Напряжение (В)	1.5 / 1.35	1.2	1.1
Максимальный объём модуля	До 8 ГБ	До 16–32 ГБ	До 64–128 ГБ
Архитектура каналов	Один 64-битный	Один 64-битный	Два 32-битных канала
Предварительная выборка	8n	8n	16n

Скорость = f (частота, тайминги, кол-во каналов)

Все должно соответствовать, и память и материнская плата и CPU.

А теперь картинки

Посмотрим, как аппаратура компонуется в серверных платформах

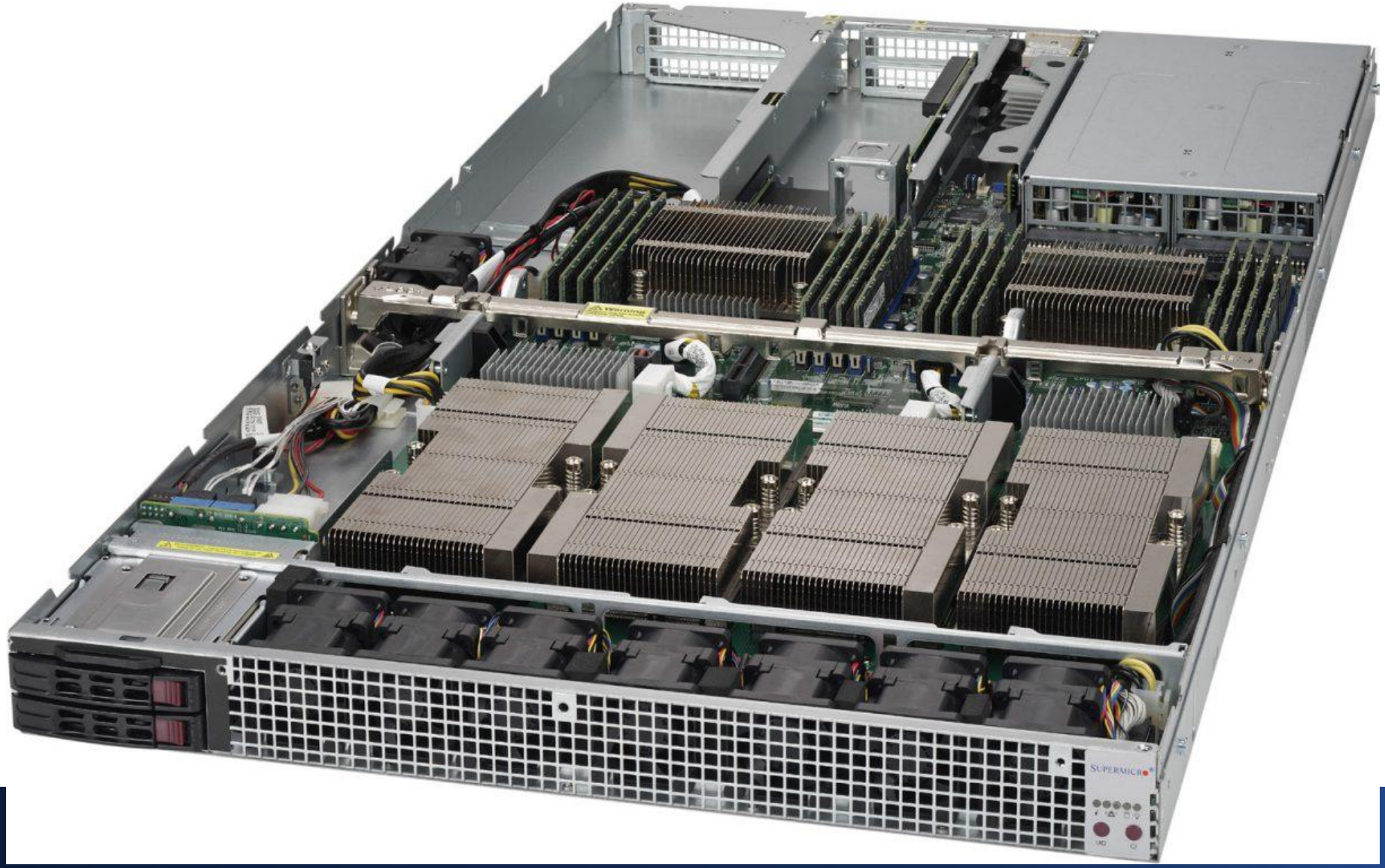


Тайминги оперативной памяти

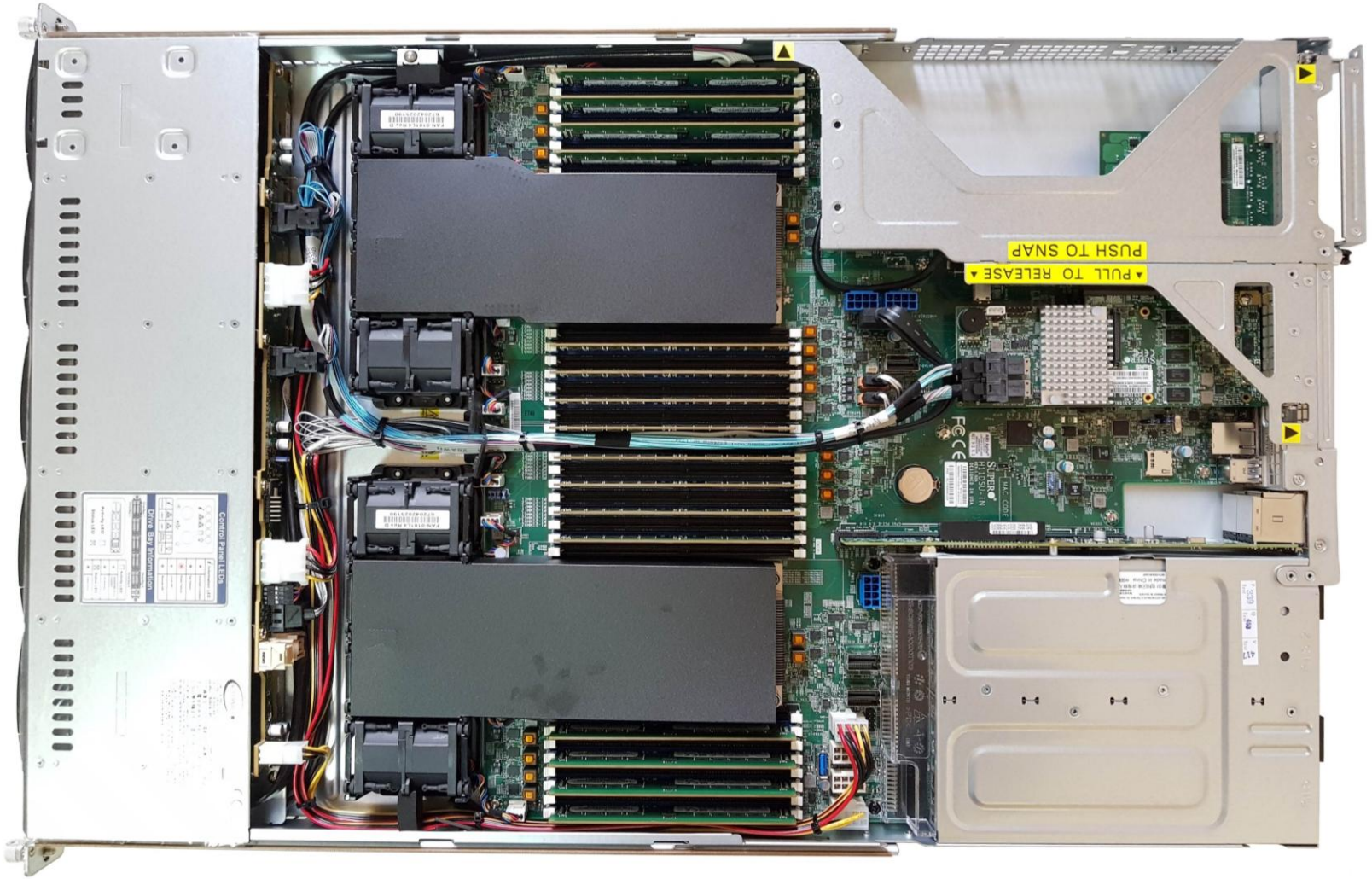
- Single системы
  - Tower
  - Rack-mount
- Блейд-системы

Форм-факторы вычислительных платформ

Форм-факторы вычислительных платформ



Форм-факторы вычислительных платформ



Красивые лампочки

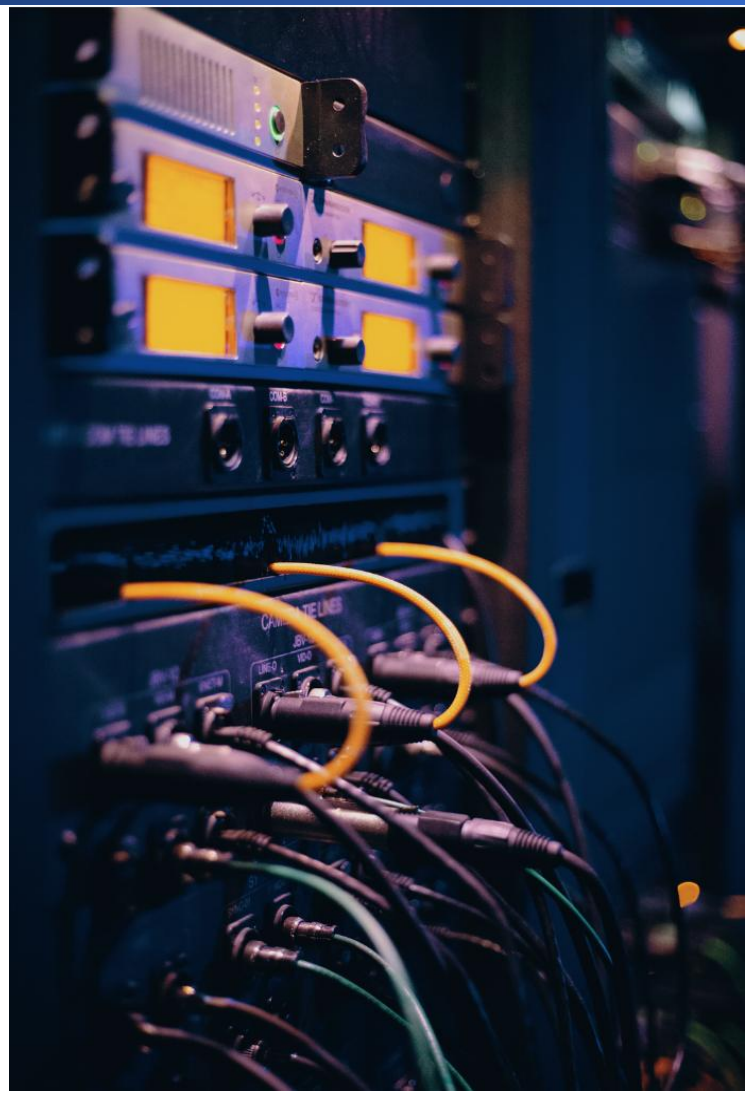
Блейд-системы

- Шасси – корпус и бэкплейн;
- Блейд-серверы (лезвия) – серверы без блоков питания, вентиляторов, сетевых разъемов и модулей управления;
- Системы питания и охлаждения для всех компонентов системы;
- Коммутационные устройства для связи с внешним миром;
- Модули управления (различные вариации на тему IPMI).

Блейд-системы

Блейд-системы

Выводы

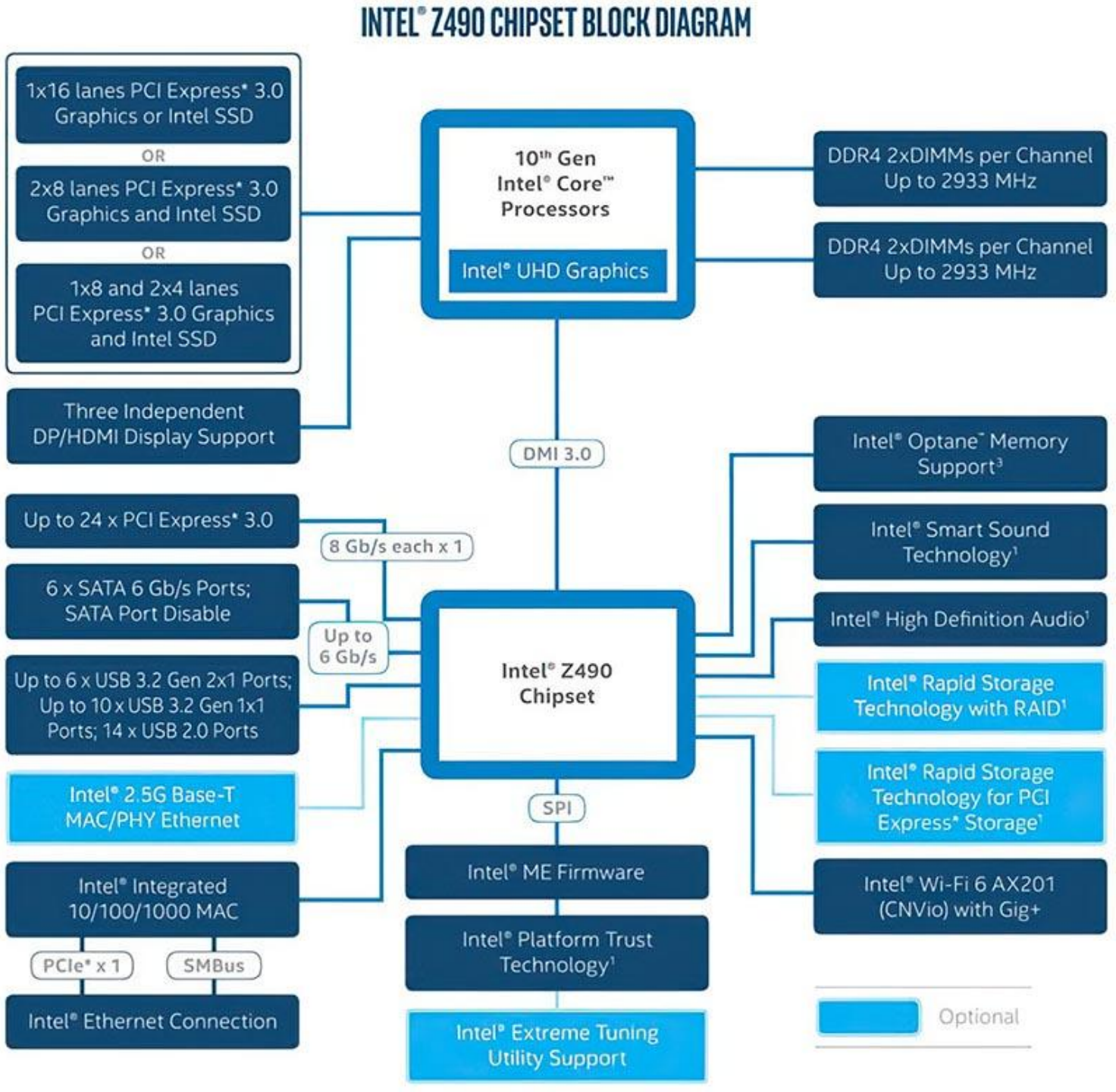


Выводы

- В современных системах используются элементы разных архитектур
- Существуют разные понятия архитектуры CPU
- У GPU своя ниша
- Основа модульной архитектуры – материнская плата
- DDR5 лучше ☺

Чипсет

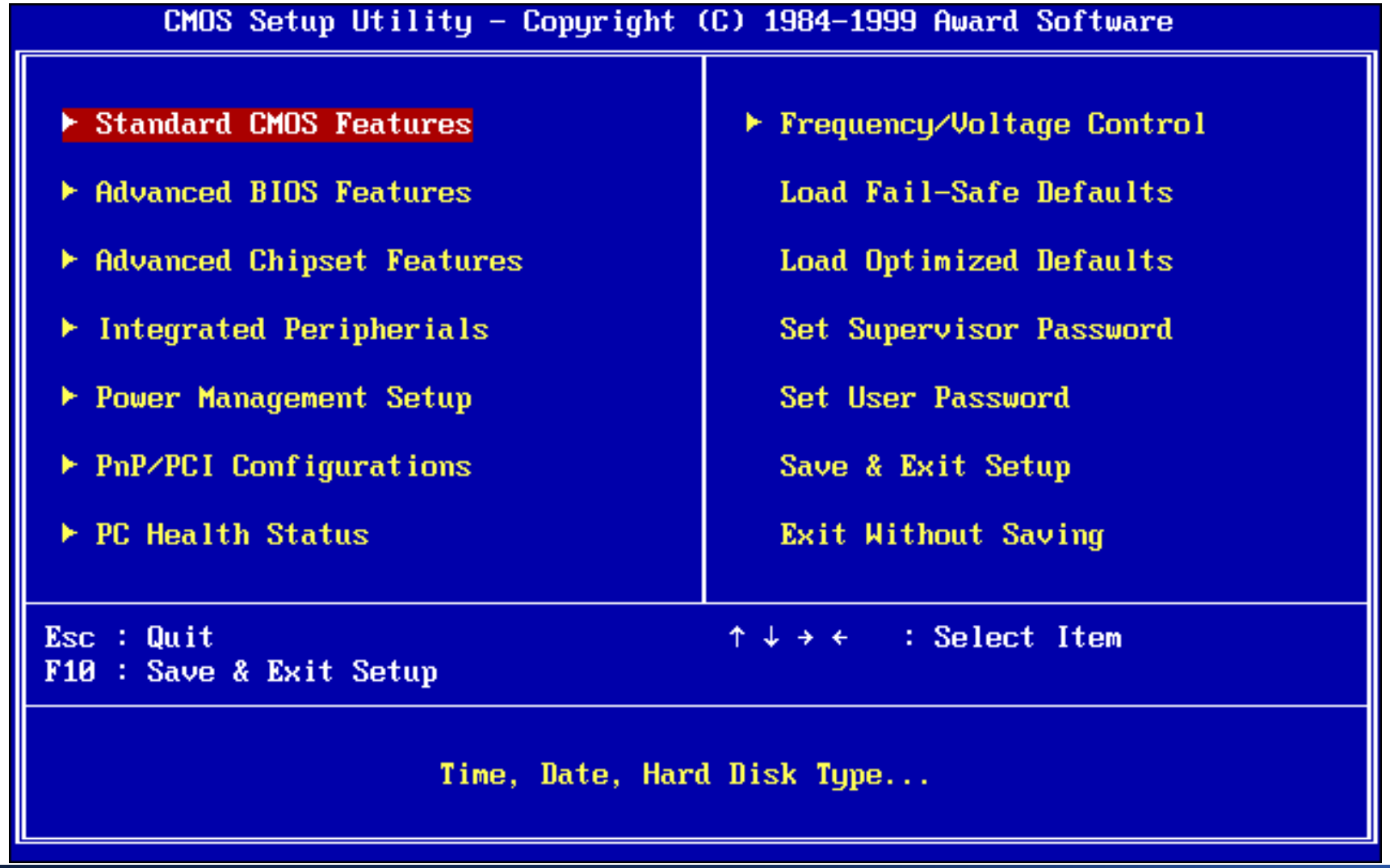
- Контроллеры системы хранения
- USB
- PXI Express
- WiFi
- Audio
- И др.



BIOS и UEFI

- BIOS - Basic Input-Output system, базовая система ввода-вывода.
  - Хранится на ПЗУ материнской платы
  - Позволяет менять параметры аппаратных компонентов
  - Обеспечивает загрузку ОС
  - BIOS загружается при включении компьютера
  - Инициализирует устройства, проводит POST (Power-On Self-Test)
- Загрузка ОС с дисков объёмом не более 2,1 Тб.
- Работает в 16-битном на 1 Мб памяти.
- Затруднена одновременная инициализация нескольких устройств

BIOS и UEFI


CMOS Setup Utility - Copyright (C) 1984-1999 Award Software
Standard CMOS Features
Advanced BIOS Features
Advanced Chipset Features
Integrated Peripherials
Power Management Setup
PnP/PCI Configurations
PC Health Status
Frequency/Voltage Control
Load Fail-Safe Defaults
Load Optimized Defaults
Set Supervisor Password
Set User Password
Save & Exit Setup
Exit Without Saving
Esc : Quit
F10 : Save & Exit Setup
↑ ↓ → ← : Select Item
Time, Date, Hard Disk Type...

BIOS и UEFI

- Unified Extensible Firmware Interface
- 2007 Intel, AMD, Microsoft и производители PC договорились о новой спецификации Unified Extensible Firmware Interface (UEFI)
- Загрузка с дисков объёмом более 2,2 Тб (до 9,4 зеттабайт).
- Может работать в 32-битном или 64-битном режимах
- Обратная совместимость с BIOS
- Безопасный запуск Secure Boot – загрузка доверенной ОС
- Доступ по сети для настройки и отладки.

BIOS и UEFI


EFI BIOS Utility - EZ Mode
Exit/Advanced Mode
06:06
Tuesday[11/16/2010]
P8P67 DELUXE
English
BIOS Version : 0304
Build Date : 10/21/2010
CPU Type : Intel(R) Core(TM) i5-2400 CPU @ 3.10GHz
Speed : 3100 MHz
Total Memory : 2048 MB (DDR3 1333MHz)
Temperature
CPU +116.6°F/+47.0°C
MB +96.8°F/+36.0°C
Voltage
CPU 1.200V
5V 5.120V
3.3V 3.408V
12V 12.288V
Fan Speed
CPU_FAN 1834RPM
PWR_FAN1 N/A
CHA_FAN1 N/A
CHA_FAN2 N/A
System Performance
Quiet
Performance
Energy Saving
Normal
Boot Priority
UEFI
UEFI
Use the mouse to drag or keyboard to navigate to decide the boot priority.
Boot Menu(F8)
Default(F5)