

# CPU相关技术

## CPU关键技术

- 微架构相关技术
- 工艺制程
- Chiplet技术、先进封装技术
- 互联技术
- 未来技术

## CPU相关组件技术

- 内存
- 硬盘
- RAID卡
- 図卡
- GPU、DPU、FPGA、ASIC
- 操作系统

## CPU微架构技术概览

■ **微架构**是遵循指令集架构的具体硬件实现方案。 **处理器微架构技术点众多**,包括执行、内存管理、并行和并发、安全、能源效率等等。在选择或者设计处理器时,这些技术需要综合考虑以满足性能、功耗、成本等多方面需求

#### 执行技术:

- 流水线执行 (Pipeline Execution)
- 指令级并行 (Instruction Level Parallelism)
- 数据级并行 (Data Level Parallelism)
- 乱序执行(Out-of-Order Execution)
- 超标量执行 (Superscalar Execution)
- 寄存器重命名 (Register Renaming)
- 矢量计算和SIMD指令 (Vector Processing and SIMD Instructions)

#### 预测和优化技术:

- 分支预测 (Branch Prediction)
- 指令预取 (Instruction Prefetching)
- 数据预取 (Data Prefetching)
- 动态频率和电压调整 (Dynamic Frequency and Voltage Scaling)

#### 并行和并发技术:

- 多线程 (Multithreading)
- 多核 (Multicore)

#### 内存和存储技术:

- 多级缓存 (Multi-Level Caching)
- 写回缓存 (Write-Back Cache)
- 写穿缓存 (Write-Through Cache)
- 内存管理单元 (Memory Management Unit)
- 虚拟内存 (Virtual Memory)

#### 可靠性和安全性技术:

- 错误检测和纠正 (Error Detection and Correction)
- 中断处理 (Interrupt Handling)
- 内存屏障 (Memory Barrier)
- 硬件锁定 (Hardware Locking)
- 硬件虚拟化支持 (Hardware Virtualization Support)
- 安全扩展,如Intel SGX (Security Extensions)

#### 功耗和热管理技术:

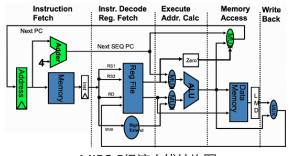
- 电源管理 (Power Management)
- 散热管理 (Thermal Management)

# **指令流水线**:空间换时间,面积换性能

■ 工业制造中,把任务细分成多个步骤,每一个生产单位只专注处理某一个片段的工作,产线像流水一样,以提高工作效率及产量,同样在处理器中采用**指令流水线技术**来提高处理器的性能



工厂流水线



MIPS 5级流水线结构图



典型MIPS 5级流水线运行图

#### 优点

- 提高吞叶量
- 可以单周期完成一条指令
- 提高CPU利用率
- 高级数CPU更高主频

#### 缺点

- 流水线冒险
  - 结构冒险(资源冲突)
    - 数据冒险 (数据顺序)
    - 控制冒险(跳转)
- 更多流水线需要更多寄存器, 需要更多面积和空间

## **分支预测**: 向左 or 向右

- 分支预测 (Branch Prediction) 用来预测下一条指令,提高CPU流水线的效率
- 当指令流水线中出现**条件分支**时,流水线中的指令需要**等待判断结果**,这就会导致了**流水线停顿**,对处理器性能有很大影响。 分支预测提前预判结果,减少流水线停顿,提高执行效率

# 向左? 向右?

#### 静态分支预测

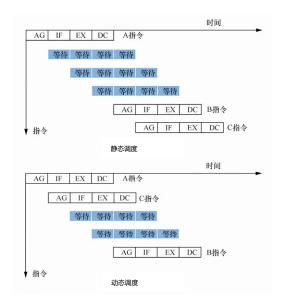
- 静态预测采用既定的策略进行预测,因此受策略本身的精确度及程序类型 影响较大
- 实现起简单、成本低,预测正确率的 波动范围很大
- 静态预测有always not taken, always taken, BFTN (backward taken, forward not taken等

#### 动态分支预测

- 动态分支预测:根据指令的不同及历 史信息作出相应的预测
- 相比静态预测,引入了分支历史的结果作为参考,预测准确率得到提高
- 动态预测策略有Saturating counter, Two-level predictor, Local branch prediction, Global branch prediction等

## **乱序执行**:高性能CPU的门槛

■ 乱序执行(Out-of-Order Execution),又称为动态执行,是指在CPU内部执行过程中,指令执行的实际顺序可能和软件中的顺序不同。根据实际情况重新排布,乱序执行极大地**提高了处理器的资源利用率和性能**。特点是"**有序**取指、重新排列执行顺序、有序结束"



乱序执行的典型例子

#### 主要步骤

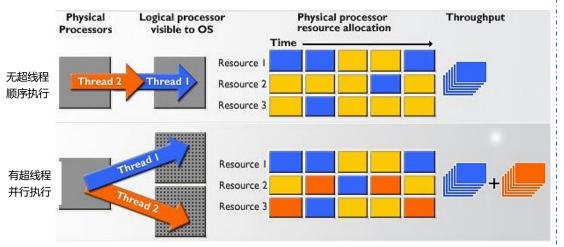


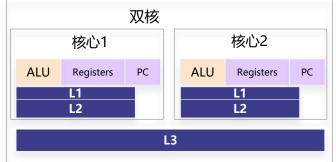
#### 优点

- 高效利用资源: 乱序执行技术允许处理器在等待数据或者资源时执行其他的指令,大大提高了资源的使用效率。
- 提升性能: 乱序执行能够让处理器尽可能地并行执行多个指令,提高了处理器的运行速度

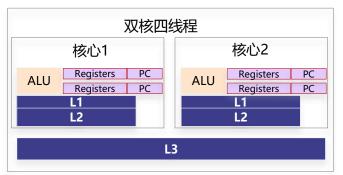
## 超线程:提高CPU利用率

超线程HT / (同时多线程SMT): 在硬件层面,它允许一个物理核心在同一时刻处理两个或更多的线程。这是通过增加一部分硬件资源(如寄存器)来实现的,当一个线程在等待内存或I/O操作时,处理器切换到另一个线程,从而提高效率





单核单线程, 双核2线程



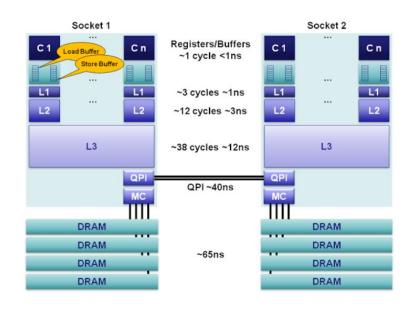
单核双线程,双核4线程

# 缓存cache: 填补CPU和内存之间性能鸿沟

■ 缓存(Cache)是为**内存**(DRAM)和**CPU**之间提供一个小但极快的存储器。CPU处理速度快,内存访问相对慢,导致等待数据导入的时间长。通过预先将经常访问的数据(指令)加载到高速缓存中,提高CPU的运行效率和性能

#### 工作原理

- · CPU缓存的工作原理主要基于局部性原理,包括时间局部性和空间局部性
- 时间局部性是指如果一个数据被访问,那么它在未来的一段时间内很可能被再次访问
- 空间局部性是指如果一个数据被访问,那么它附近的数据在未来的一段时间内很可能被访问





#### 缓存分级

缓存级别	特点
L1	最接近CPU,速度最快,容量最小, 包含指令缓存和数据缓存
L2	离CPU更远,速度稍慢,容量较大
L3	离CPU最远,速度最慢,容量最大

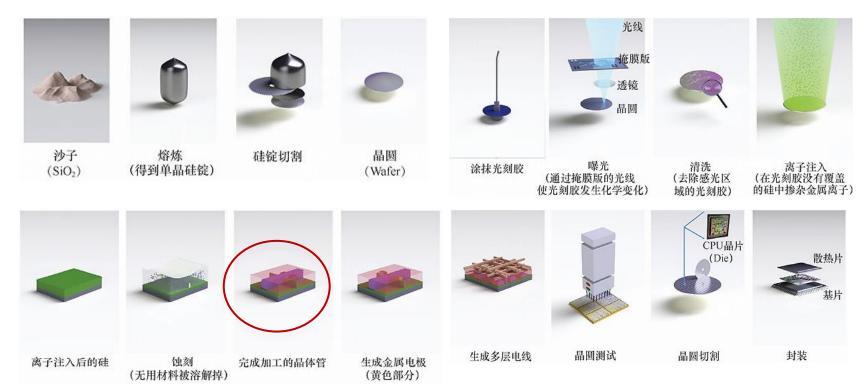
L1、L2、L3,是固定晶体管数目达成平衡的综合效果

#### 缓存一致性

- 在多核或者多个处理器系统中,**各级Cache和内 存池可能对同一份数据有多份副本**,如何保证这 些副本的一致性(Coherency)是个难题
- 需要Cache一致性协议来保证,如Intel使用 MESIF协议,AMD使用MOESI协议等

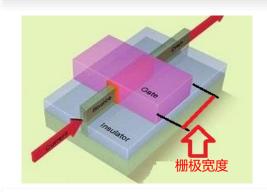
# 工艺制程: CPU生产过程 从沙子到CPU

■ 完整生产流程有几百道工序, 生产设备都是尖端制造工具。3种基本手法: **生长、挖掉、掺杂** 



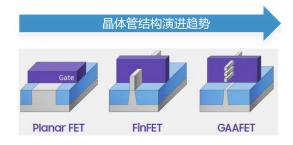
## **工艺制程**: 指甲空间 建繁华城市

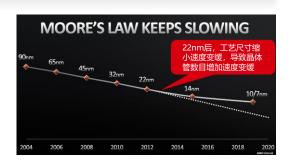
■ CPU的**纳米工艺**的通俗定义是指"栅极沟道的最小宽度"。一个晶体管有3个引脚,晶体管导通时,电流从源极(Source)流入 漏极(Drain),中间的栅极(Gate)负责控制源极和漏极之间电流的通断,栅极的最小宽度就是纳米工艺中的数值



- 2000年后, xx 纳米 与晶体管的任何实际物理特征 (栅极长度、金属间距或栅极间距) 无关。
- 根据 IEEE协会数据3 纳米节点的接触栅极间距为 48nm, 最紧密的金属间距为 24 nm。
- 最近的 "3 nm" 主要被芯片制造商用作营销术语, 指代在晶体管密度增加、提高速度、降低功耗

https://en.wikipedia.org/wiki/3 nm process





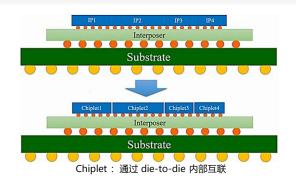
				工女		中小士 17	$\pi$		
1圆厂	2015	2016	2017	2018	2019	2020	2021	2022	2

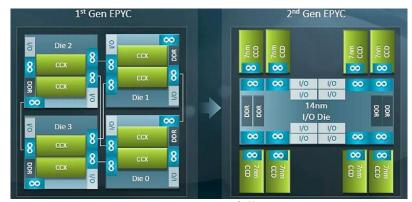
晶圆厂	2015 2016	2017 2018	2019 2020 2021	2022 2023	2024 2025
台积电	16nm	10nm 7nm	7nm+ 5nm	3nm	2nm
三星	14nm	10nm 8nm	7nm 5nm	3nm	
英特尔	14nm 14nr	n+ 14nm++	10nm 10nm+ 7nn	n 7nm+ 7nm+	+
格罗方德	14nm	12nm			
联电	28nm	14nm			
中芯国际	28nm		14nm		

**土田日回口制担共占** 

# Chiplet技术: 延续摩尔定律的关键技术之一

- 随着摩尔定律放缓,芯片设计成本、时间、性能等面临严峻挑战,chiplet技术应运而生。
- Chiplet技术它将不同的芯片 (chiplet) 组合在一起,形成一个更大、更复杂的系统。每个chiplet可以专门用于执行特定的功能,例如处理器核心、内存、I/O接口等。这些chiplet可以通过高速互连技术(如Intel的EMIB或AMD的Infinity Fabric)连接在一起,形成一个统一的系统芯片
- 灵活性和可扩展性、成本效益、性能优化、创新加速





AMD ZEN1、ZEN2架构

	单片	Chiplet芯片
设计成本	高,7nm大于2亿美元	比单片成本低
设计时间	厂,一般大于18个月	较短,一般12个月
设计风险	高,遗漏功能需要重新设计	较低, 重新设计内容, 可增减模块芯片
功耗	低	接近单片
上市时间	慢	较快
产品尺寸	小	较小

- 2022年3月,英特尔、AMD等芯片厂商与Google、微软等科技巨头共同成立了Chiplet标准联盟
- 研究机构Omdia的报告显示,到2024年,采用Chiplet处理器芯片的全球市场规模将达58亿美元,到2035年,有望达到570亿美元

## CPU封装技术: 持续提高集成度

- 封装 (packaging, PKG) : 利用膜技术及微细连接技术,将半导体元器件及其他构成要素在框架或基板上布置、固定及连接,引出接线端子,并通过塑性绝缘介质灌封固定,构成整体主体结构的工艺
- CPU优良的封装技术能提供更好的**散热**、更优**电气特性**、更强的**机械强度**,以满足日益严苛的性能需求。

## 传统封装

通孔插装、表面贴、球栅阵列:引线框架作为载体,引线键合互连

• 平面网格阵列封装 (LGA)



触点都在CPU的PCB上,针脚在主板 Intel自775之后的所有桌面处理器 AMD 皓龙、霄龙、TR等处理器

• 插针网格阵列封装 (PGA)



针脚在了CPU的PCB身上,插孔在主板 intel 775以前的大部分桌面处理器 AMD 几乎全部的家用桌面处理器 intel 大部分以M,MQ结尾的移动处理器

• 球柵网格阵列封装 (BGA)

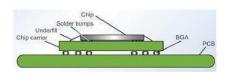


引脚以球形排列在芯片底部 intel所有以H/HQ/U/Y等结尾低压处理器 AMD 低压移动处理器 所有手机处理器

#### 先讲封装

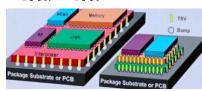
多芯片组装、立体结构型: I/O 数量多、芯片相对小、高集成化

• 倒装芯片结构封装 (Flip Chip)



Flip Chip: 芯片上制作金属凸点,芯片面朝下利用焊料与基板互连,底部填充树脂 封装面积小,引线互连长度缩短,I/O端口数 量增加

• 2.5D封装、3D封装



2.5D封装 3D封装

2.5D封装: 芯片和封装载体间加入硅中介转接层,利用硅通孔 (TSV) 连接其上、下表面的金属,多采用倒装芯片组装工艺容量、性能比2D提升巨大

3D封装: 芯片与芯片直接堆叠,可采用引线键合、倒装芯片或二者混合的组装工艺及TSV技术进行互联缩小产品尺寸,提高了产品容量和件能

## CPU互联技术

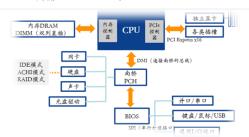
## 当前互联技术

CPU互联总线协议用于处理器各个部件之间的 通信,随着处理器架构的不同而不同

- **片内总线**:通常是专有的,提供了高带宽、低延迟的点对点连接。
  - X86: DMI、QPI、UPI; UMI、HT、Infinity Fabric:
  - ARM: AMBA系列总线: AHB、APB、AXI
- 片外总线:通常是标准的,实现与外部设备的数据 交换和通信

• 内存: DDR4/DDR5; 硬盘: SATA

• 外部IO: PCIE4.0, PCIE5.0



举例: Intel CPU内部互联

## 发展中互联技术

**UCle**: Chiplet早期多公司开发专有接口的互连,以实现自家芯片间的互连;通过发展,当前UCle最有希望成为业界统一D-D的互联标准

Standard	owner	Throughput (Gbps)	Latency	Package Types
AIB+	Intel	2	<5ns	Bridge
HBM	JEDEC	6.4		2.5D
Open-HBI	OCP-ODSA	8		2D,2.5D
Lipincon	TSMC	2.8	<14ns	2.5D
BOW	OCP-ODSA	16	<5ns	2D,2.5D
Infinity Fabric	AMD	16	<9ns	2D
UCle	Intel	32	<5ns	先进封装: 2.5D、桥接 标准封装: 有机衬底或层压板

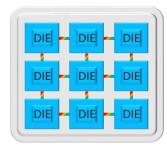
**CXL**: 允许CPU和外设直接访问彼此内存,降低延时, 提高性能。用于内存扩展、内存池化、内存共享等应用



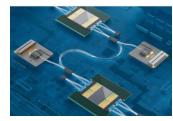


## 未来互联技术

**光I/O互联**:具有高带宽、低延迟及低功耗等优点,可实现新型异构、可组合计算系统架构



芯片内光互连

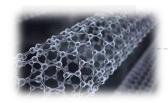


芯片间光互连

# 后摩尔时代, CPU未来技术突破

随着摩尔定律的逐渐变缓,新材料的应用,新原理,新信息载体等创新技术不断突破,以制造占用**面积更小、计算能力更强、能耗更低**的处理器芯片

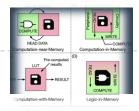
#### 新材料应用



#### 碳基处理器

- 更优的性能和更低的功耗
- 制程工艺要求不高
- 北京大学团队、MIT团队研究
- 2021年,华为发布了石墨烯晶体管专利

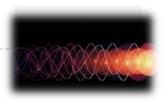
#### 新原理



#### 存算一体处理器

- 计算单元与存储单元融合,**同时** 完成数据存储和计算功能
- 解决冯·诺伊曼瓶颈,做到低成本、 大算力的同时,保持低功耗
- 国内10来家初创: 干芯科技、后 摩智能、知存科技...

#### 新信息载体



#### 光子计算处理器 量子计算处理器

- 光子计算利用光子来传输和处理信息:超高速度、超强的并行性、超高带宽、超低损耗
- 量子计算是一种利用量子力学原理 处理信息的计算方式: 更高的速度 并行处理大规模的数据



# CPU相关技术

## CPU关键技术

- 微架构相关技术
- 工艺制程
- Chiplet技术、先进封装技术
- 互联技术
- 未来技术

## CPU相关组件技术

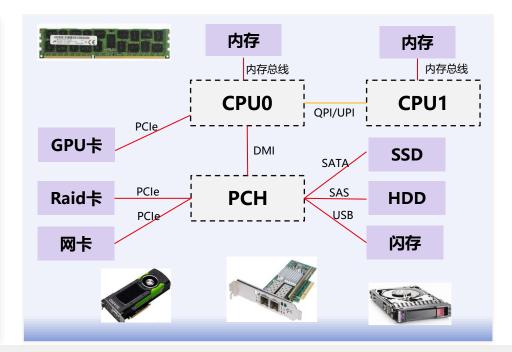
- 内存
- 硬盘
- RAID卡
- 网卡
- GPU、DPU、FPGA、ASIC
- 操作系统

## CPU外围组件概览

"CPU+芯片组+总线"构成不同的 CPU "平台"系统,外围组件主要包括:内存、硬盘、 RAID卡、

## **网卡、DPU、GPU及其他AI加速器**等

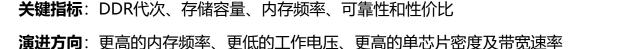
- 芯片组PCH: 平台控制器集线器,处理输入输出 (I/O)数据,负责 PCIe 总线、USB、SATA、音 频控制器、键盘控制器等间的通信。
- 通过PCIe 总线扩展 I/O 设备,典型如GPU卡、
   DPU卡、其他AI 加速卡、PCIe 网卡等
- 系统数据传输过程: 数据经由网卡封装与解封、链路管理、数据编码与 译码后,储存于硬盘中;当程序需要执行时,将数 据从硬盘经由内存至 CPU。



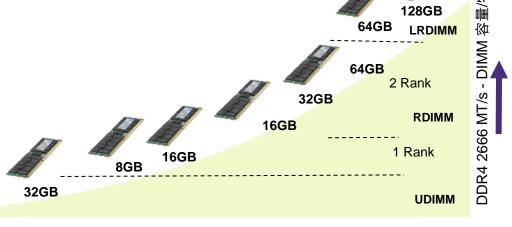
# 内存: CPU与外部存储沟通桥梁

内存是计算机中重要的部件之一,其作用是**暂时存放CPU中的运算数据,以及硬盘等外部存储器交换的数据。**计算机中所有程序的运行都是在内存中进行,内存的性能对计算机的影响非常大。





代次	量产时 间	时钟频率 MT/s	工作电压V	引脚数量 PIN
DDR	2002	200-400	2.5	184
DDR2	2006	400-800	1.8	240
DDR3	2010	1066-1866	1.5	240
DDR4	2012	2133-3200	1.2	288
DDR5	2020	4800-6400	1.1	288



## 硬盘:数据的归宿



Platters Spindle

R/W Head

ctuator Arm

ctuator Axis Actuator

机械硬盘

固态盘

Cache Controller NAND Flash Memor

机械硬盘 (HDD): 使用旋转的磁性介质来存储数据。可通过执行器臂上的读/写头访问数据。 **固态盘** (SSD): 使用基于闪存的存储器来存储数据。具有比机械硬盘**更快的读/写速度**、更 短的访问时间以及更高的每 GB 存储成本。

关键指标: 硬盘容量、转速、平均访问时间、传输速率、IOPS等

项目	固态硬盘	传统硬盘
容量	较小	大
价格	盲	低
随机存取	快	一般
写入次数	SLC:10万次 MLC:1万次	无限制
工作噪声	无	有
防震	好	差
数据恢复	难	可以
重量	轻	重

#### 2.5in NVMe SSD



#### M.2的固态硬盘



## RAID卡: 提升性能 守护数据安全



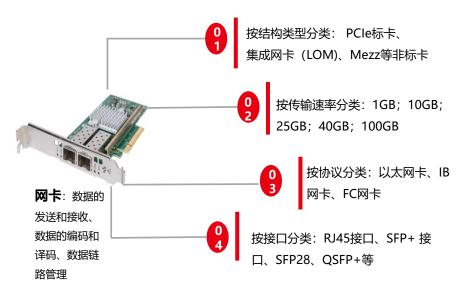
RAID 数据安全保护功能	更有效的数据组织	
---------------	----------	--

RAID级别	RAID 0	RAID 1	RAID 3	RAID 5	RAID10
容错性	无	有	有	有	有
冗余类型	无	镜像冗余	校验冗余	校验冗余	镜像冗余
可用空间	100%	50%	(N-1)/N	(N-1)/N	50%
读性能	高	低	高	高	普通
随机写性能	高	低	低	低	普通
连续写性能	高	低	低	低	普通
最少磁盘数	2	2	3	3	4
应用场景	传输带宽需 求大的应用	安全性要求 较高的应用	大文件、连 续数据应用	读/写比率 较高的应用	安全性要求 高的应用





## 网卡:连接网络通信



智能网卡:除了具备标准网卡的功能外,主要实现网络业务卸载。智能网卡上没有CPU,需要Host CPU进行管理。



- Host offload
- ✓ virtual switching (OVS),
- √ software-defined storage
- data and network encryption (IPSEC/SSL)
- NFV
- Security Isolation



支持TCP/IP协议,应用于以太网络中



特指Infiniband网卡,即IB卡。应用于高带宽,低时延的高性能计算机中



特指FC网卡,支持FC协议, 连接存储或光纤交换机



在机箱内部安装



夹层卡,比标准的PCle连接 器小,节省空间

## ▮ DPU: 灵活硬件卸载能力

DPU (Data Processing Unit) 是以数据为中心构造的专用处理器,采用软件定义技术路线支

- 第一阶段:基础功能网卡
- 基础功能网卡(即普通网卡)提供2x10G或2x25G带宽吞吐
- 具有较少的硬件卸载能力
- 支持SR-IOV. 支持多队列
- 第二阶段:智能网卡
- 智能网卡以**数据面卸载**为主
- 较丰富的硬件卸载能力 (OVS Fastpath)
- RDMA网络卸载 (RoCEv1、RoCEv2)
- 无损网络能力 (PFC、ECN、ETS等)
- NVMe-of的硬件卸载
- 安全传输的数据而卸载
- 第三阶段: DPU
- 运营商领域对智能网卡与DPU称呼不做区分。
- 业界一般认为可以认为: **DPU= CPU + 智能网卡**。
- 卸载控制平面的任务和一些灵活复杂的数据平面任务



DPU智能网卡

## GPU: 拥有超级并行计算能力



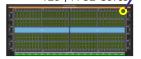
- · AI应用具有计算密集、矩阵运算、极高访存需求的特征
- 智算AI系统面对多个场景,每个场景用户需求五花八门,多个场景叠加成了多样的差别化的用户需求。**所以通用性是首先要考虑的因素。**
- · GPU具有灵活可编程能力和AI算法的需求是相对匹配的,是目前AI算力的主流平台。

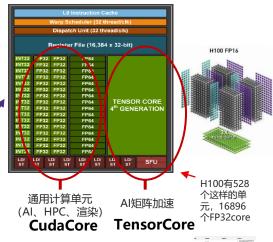
#### GH100芯片有144 SMs:

- H100 SXM5: 132个 SM (16896个 FP32core)
- H100 PCle5: 114个 SM (14592个 FP32core)

#### 每个SM处理器:

• 128个FP32 Cores





**Nvidia GPU** 



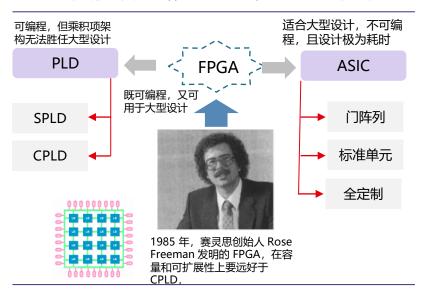
**优势**:训练速度快、通用性强 **劣势**:功耗较高、硬件不可编辑

应用场景: GPU广泛应用于PC,游戏主机,汽车,AI服务器及移动端领域

## FPGA: 能 "经常改" 、 "算得快"

## FPGA:现场可编程门阵列,可由最终用户配置实现许多复杂的逻辑功能。

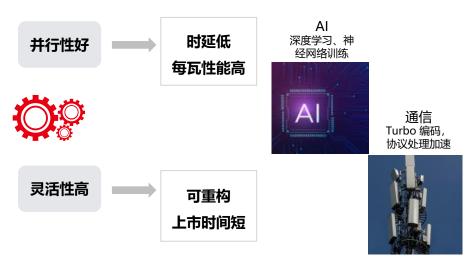
## FPGA 发明初衷是弥补当时 PLD 和 ASIC 之间的缺环



**关键指标**:制程、逻辑资源、IO数、固化单元、功耗

**演进方向**:逻辑单元数不断提升,集成度不断提高,向**异构化**发展。

FPGA 灵活性高、并行性好,特别适合小批量、时延要求低、变化快的场景



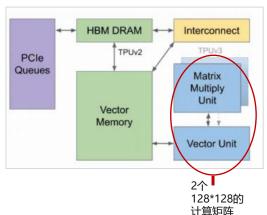
应用场景: AI、通信(防火墙)、工业、汽车、安防、航天、芯片...各细分行业

# ASIC: 特定场景高效的AI加速

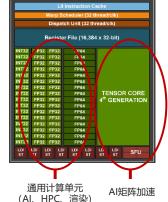
ASIC: 即专用集成电路,指应特定用户要求和特定电子系统的需要而设计、制造的集成电路

谷歌的TPU、寒武纪的MLU,网络处理器都属于ASIC芯片.

举例: TPU专注于矩阵运算, 牺牲处理器的通用性



仅仅支持神经网络所需的大规模的加乘运算 提升了神经网络的运算效率,节省了功耗



CudaCore

TensorCore

Nvidia GPU 部分SM



优势:面积小、效能高、功耗低、专用定制

**劣势**: 开发周期长、硬件可编程差

目前国内AI芯片厂商有70+家,多数是面向具体场景的,如语音、人脸、安防、自动驾驶等细分方向













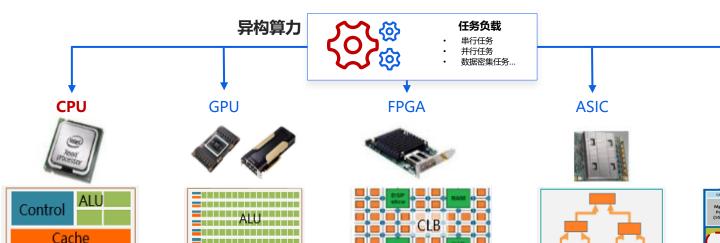








# CPU和其它专用处理器 共同构建异构算力



• 计算单元、控制单元,存储单元

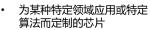
DRAM

- 大部分晶体管用于构建控制电路和高速内存,计算单元少(30%)
- 用于管理和调度任务
- · X86: Intel,AMD; ARM: 华为,飞腾...

• 计算单元、控制单元,存储单元

DRAM

- · 大量计算单元 (80%) 和高速 内存,逻辑控制单元简单
- 擅长做3D图像处理、并行计算、 AI计算
- NVIDIA (GPU); AMD(APU)
- 可编程逻辑,在芯片内集成 大量的基本门电路,存储器 及互连资源。
- 适合算法更新频繁及协议解析处理场景:视频,基因,网络等应用
- Intel(Altera);AMD(Xilinx)



- 应用在专用领域,性能高、 灵活性差
- 寒武纪思元,谷歌TPU,海思 Ascend ,Intel VPU(Movidius)



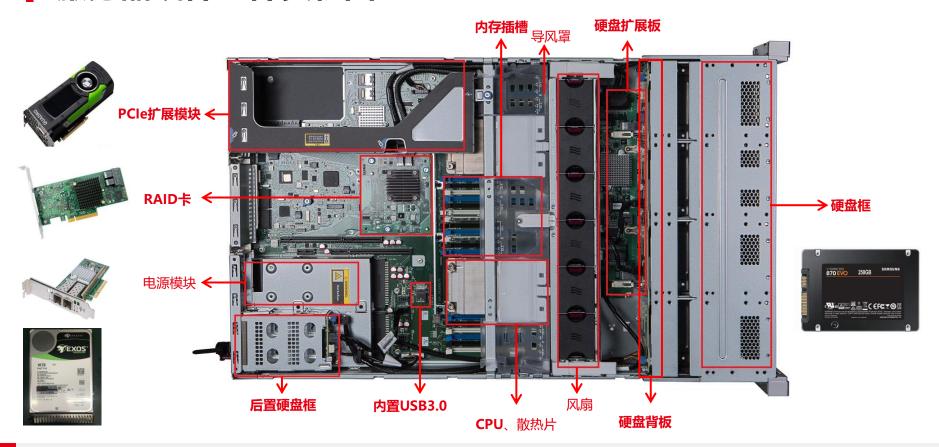
**DPU** 



- DPU把ARM处理器核、VLIW矢量计算引擎和智能网卡的功能集成在了一起,形成SOC
- 主要应用在卸载CPU任务,用 于分布式存储、网络计算和网络 安全应用
- NVIDIA、Intel、博通

GPU,FPGA,ASIC,DPU各自具有独特特长和应用领域

# 服务器硬件组件实物图



## BIOS和操作系统

■ BIOS相对比较封闭,与CPU厂商紧密相关。服务器使用的BIOS包括:AMI、百敖、中电昆仑,不同的BIOS对于不同的机型。

■ 主流服务器系统Windows Server,Redhat、SUSE、 Euler四大体系;<mark>麒麟</mark>在国产商用操作系统占据主导地位, 其次是统信的<mark>UOS</mark>,两种操作系统占国内<mark>信创市场95%</mark>以上。Open Euler分为社区版和LTS版,华为开源的目标主要是带动服务器销售和发展国产芯片|操作系统|系统软件产业链整体发展。我司服务器OS除了支持NingOS外,还支持麒麟、统信UOS、欧拉等,不同的机型有不同的OS支持列表

#### **BIOS**

带内管理, 初始化硬件, 完成硬件枚举和自检, 启动操作系统

#### 操作系统

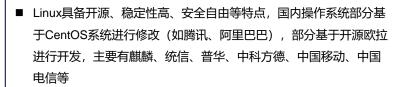
对计算机或服务器进行资源管理,为用户提供方便的操作界面 核心功能:进程调度、内存管理、文件系统、设备管理、用户界面、 网络协议



国产化信创领域



- 遵循UEFI标准,当前X86全球3家: AMI、Insyde、百敖
- 国产ARM: 华为选百敖, 飞腾选中电昆仑, 海光用百敖



	麒麟	统信	普华	NingOS	欧拉
产品类型	桌面OS,服务器 OS	桌面OS,服务 器OS	桌面OS,服务器 OS	服务器OS	鲲鹏ARM服务器
内核	Linux	Linux	Linux	Linux	Linux
厂家	中国电子CEC	诚迈科技/统信	中国电子CEC/太 极股份	新华三	华为
适配CPU	所有国产CPU	所有国产CPU	X86, open Power, 龙芯, 申 威, 兆芯	X86、ARM所 有H3C服务器	鲲鹏CPU
商业版本	Y	Y	Y	N,公司内所 有使用服务器 操作系统的产 品	N , 开源且不做商业版。推测主要是带动 其芯片业务,发展鲲鹏 生态。