CPU发展报告

CPU（Central Processing Unit），即中央处理器，是计算机系统的运算和控制核心。1971年，Intel公司推出世界上第一款微处理器4004；2022年，Intel公司推出12代酷睿处理器。五十年间，随着集成电路的发展，CPU发展迅速。

（一）CPU架构发展

CPU架构是CPU厂商给属于同一系列的CPU产品定的一个规范，主要目的是为了区分不同类型CPU的重要标示。目前市面上的CPU分类主要分有两大阵营，一个是Intel、AMD为首的复杂指令集CPU，另一个是以IBM、ARM为首的精简指令集CPU。

X86架构：1978年6月8日，Intel发布了史诗级的CPU处理器8086，由此X86架构传奇正式拉开帷幕。首次为8086引入X86作为计算机语言的指令集，定义了一些基本使用规则，X86架构使用的是CISC复杂指令集。同时8086处理器的大获成功也直接让Intel成为了CPU巨头，我们所熟悉的奔腾处理器、酷睿处理器使用的都是X86架构。

MIPS架构：在上世纪80年代由美国斯坦福大学Hennessy教授的研究小组研发，它采用精简指令系统计算结构（RISC）来设计芯片。和Intel采用的复杂指令系统计算结构（CISC）相比，RISC具有设计更简单、设计周期更短等优点，并可以应用更多先进的技术，开发更快的下一代处理器。MIPS是出现最早的商业RISC架构芯片之一，新的架构集成了所有原来MIPS指令集，并增加了许多更强大的功能。MIPS自己只进行CPU的设计，之后把设计方案授权给客户，使得客户能够制造出高性能的CPU。2007年，中科院计算机研究所的龙芯处理器获得了MIPS的全部专利、指令集授权，中国开始走上了以MIPS为基础的CPU研发道路。

PowerPC架构：PowerPC是由蓝色巨人IBM联合苹果、摩托罗拉公司研发的一种基于RISC精简指令集的CPU，PowerPC架构最大优点是灵活性非常好，核心数目灵活可变，因此在嵌入式设备上具有很高效益，可以针对服务器市场做超多核，针对掌机做双核，因此它具有优异的性能、较低的能量损耗以及较低的散热量。

ARM架构：ARM可以说是一个异军突起的CPU架构，采用了RISC精简指令集，而且ARM发展到今天，架构上非常灵活，可以根据面向应用场景不同使用不同设计的内核，因此可以广泛用于嵌入式系统中，同时它高度节能的特性，目前各种移动设备中全都是它的身影。据统计，使用ARM架构的芯片年出货量高达200亿片，随着物联网时代降临，对于低功耗性ARM芯片需求量会发生爆炸性增长。

虽然历史上出现过许多的CPU架构，它们之间的差异很大。但是，经过时间和用户的检验，现在主流的CPU架构只剩下X86架构和ARM架构。二者区别主要在于采用的指令集不同。前者采用CISC复杂指令集，而后者采用RISC精简指令集。CISC可以实现高性能CPU设计，但是设计起来相当麻烦，要保持庞大硬件设计正确是一件不容易的事情，还要确保性能有所提升，不能开倒车，因此桌面CPU研发时间慢慢变长。这时候，以ARM为首的一些RISC精简指令系统计算机开始崭露头角了。

对于传统的X86处理器，如果是多核，内部的多个核心都是一模一样的，这样的话，由于一旦软件只能调度一个核心，处于高频工作，但由于架构限制，其余核心也要保持同样的高频率和高电压状态，这样就浪费了大量的能量在做无用功。后来就发展出了异步多核，允许不同核心工作在不同频率上，以此换来更低功耗。

由于移动设备更加在意功耗，所以ARM采用了更加激进的做法，八个核里面允许有不同Cortex-A架构核心，那就是著名的ARM bigLITTLE。这样的大小核设计目的很明确，就是在有限的电池容量中，兼顾性能、续航的需求，因此SoC内部的CPU是采用异构计算，既有高性能大核心，也有低功耗小核心。这种大小核架构后来的Intel酷睿处理器也开始采用。

（二）CPU发展

1. 4004处理器：

1971年，Intel生产出世界上第一块四位CPU，代号4004。基于P沟道MOS硅栅技术，采取10μm工艺，拥有2250个晶体管，主频740KHz，处理速度达到0.06 MIPS，前端总线为0.74MHz（4位），socket为DIP16。4004处理器可以执行4位运算，支持8位指令集和12位寻址，指令集为4位BCD编码。其大小比人类手指甲还小，但是运算性能可以与ENIAC媲美。

优点：第一款商用微处理器，相对于ENIAC，体积小，运算效率高。

2. 8008处理器：1972年，Intel生产出世界上第一块八位CPU，代号8008，也是首款在PC上使用的处理器。基于P沟道MOS硅栅技术，采用10μm工艺，拥有3500个晶体管，主频0.5MHz，处理速度达到0.05MIPS（略慢于4004），socket为DIP18。8008处理器可以执行8位运算，支持14位寻址，可以访问到16KB的内存。虽然其单位时间执行指令数稍慢于4004，但由于其8位运算的优势使其在大多数应用中拥有更好的性能表现。

改进：晶体管数目增加，主频变慢，支持8位运算。

缺点：相较于4004处理器，单位时间执行指令数较少。

优点：可以执行8位运算，大多数应用中性能表现较4004处理器提升明显。

3. 8080处理器：1972年，Intel发布8080芯片，依然为8位处理器，以N沟道MOS取代了P沟道MOS，使用6μm工艺，拥有6000个晶体管，主频2MHz，处理速度达到0.64MIPS，运算速度比8008提升了10余倍，socket为DIP40。8080处理器在8008处理器的基础上进行了改进，其地址总线为16位，可以访问64KB的内存，是8008的四倍；NMOS取代PMOS的升级；其指令集比8008指令集进一步增强等。8080的成功使其直接影响了后续的X86架构处理器。

改进：以N沟道MOS取代了P沟道MOS，使用6μm工艺，晶体管数目增加，主频加快，地址总线宽度增加到16位，指令集相比8008指令集进一步增强。

优点：运算速度比8008提升了10余倍，可以访问64KB的内存，是8008的四倍，指令集相比8008指令集进一步增强。

4. 8086/8088处理器：1978年，Intel推出微处理器8086，也是第一款16位处理器，采用3μm工艺与特色的x86-16指令集，拥有2.9W个晶体管，主频4.77MHz，处理速度达到0.33MIPS，socket还是DIP40不变。之后，Intel公司发布了8088处理器，采用x86指令集，实际上是外部数据总线削减为8位的8086。

改进：采用3μm工艺与特色的x86-16指令集，晶体管数目增加，数据总线改为16位

缺点：在8086中，对单一指令仅仅采用2级步骤完成（取指，执行）同时CPU内只能执行一条指令，非流水线操作。在实模式下用户程序和操作系统拥有同等权利，对访问的地址即为物理地址，而同时程序可以随意修改自己的段基址，因此可以随意修改任意物理地址甚至操作系统的内容；同时，因为各程序使用的逻辑地址即物理地址——“所见即所得”——故微处理器无法支持”多任务“功能。

优点：主频加快，处理速度提升，后来被作为X86架构的鼻祖，其使用的X86指令集不断被使用、拓展至今。

5. 80286处理器：1982年，Intel发布了16位处理器80286。其兼容8086的所有功能，采用1.5μm工艺与相同的x86-16指令集，拥有13.4W个晶体管，主频6MHz与外频相同，处理速度达到0.9 MIPS，socket为正方形包装的PGA68。

改进：采用1.5μm工艺，晶体管数目增加。

缺点：在80286中，对单一指令采用3级步骤完成（取指，译码，执行）但是CPU内只能执行一条指令，非流水线操作。

优点：其内、外部数据总线皆为16位，地址总线24位，可寻址16MB内存。新增保护模式，改进了实模式下内存访问的不安全性。

6. 80386处理器：1985年，Intel发布了第一款32位处理器80386，采用相同的1.5μm工艺与扩充后的x86-32指令集，拥有27.5W个晶体管，主频12.5 MHz与外频相同，处理速度达到6 MIPS，性能比80286提升约10倍，socket为PGA132。

改进：晶体管数目增加，增加数据总线和地址总线的宽度，增加虚拟8086模式，引入L1 Cache，添加内存管理单元，扩充指令集内容。

缺点：在80386中，对单一指令依然采用3级步骤完成（取指，译码，执行），非流水线操作。

优点：其数据总线以及地址总线皆为32位，因此能够寻址4GB的内存空间，因此为保护模式的优化创立条件——虚拟8086模式。引入了新的处理器工作模式——虚拟8086模式，真正实现各色程序的多任务处理功能，能兼容之前的老机型程序。首次引入了L1 Cache（片外），极大的提高了CPU访问内存的效率。添加了内存管理单元（MMU），从而使得处理器能够有内存保护，存储器分页管理机制等功能。

7. 80486处理器：1989年，Intel发布了32位处理器80486，采用1μm工艺，指令集为x86与x87的集合（增加了浮点运算部件），拥有120W个晶体管（突破百万晶体管的界限），主频25 MHz，处理速度达到20 MIPS，socket为PGA。

改进：引入流水线的概念，采用RISC技术，引入Forwarding技术和Pre-fetch技术。片内集成了指令数据统一的8KB L1 Cache，同时支持外部L2 Cache。片内集成了增强型FPU单元，提出时钟倍频技术，改进MMU的性能，扩充指令集，采用新的内部总线传送技术。

优点：加快指令执行效率，实现一个时钟周期内完成一条指令。解决了数据冒险问题，分支操作加快。浮点计算性能增强，CPU主频增加，大大提高了与内存的数据交换速度。

8. Pentium处理器：1993年，Intel发布了32位处理器Pentium，又被号为80586，使用的超标量X86微体系架构（称为P5架构），是8086兼容处理器系列中的第五代主要产品，指令集非常接近80486与80386，仅仅做了简单的指令内容增加。其采用0.8μm工艺，拥有310W个晶体管，主频60MHz，处理速度达到100MIPS（8048的四倍性能），插槽采用Socket4/5/6。

改进：采用0.8μm工艺，晶体管数目增加，采用超标量技术。引入对L1 Cache进行数据指令分离，引入了MESI的Cache coherence协议。

缺点：L2 Cache依然是片外的，极少数情况下除法运算的精度低。Pentium依然采取in-order执行（指令）；同时CPU面对data hazard时pipeline中依然会出现stall。

优点：系统运行效率大大提升，浮点性能提升，实现更高效的cache写回，解决了发热问题。

9. Pentium MMX处理器：1996年，Intel推出了基于P5架构的微处理器Pentium MMX，添加了MMX指令集，是对P5架构的一次关键升级。采用0.35μm工艺，拥有450W个晶体管，主频166MHz，插槽采用Socket7.

改进：添加了Intel的MMX (Multi-Media Extension)指令集，L1 Cache升级为双倍大小。

优点：处理多媒体的能力上提高约60%

10. Pentium Pro处理器：1996年，Intel发布了32位处理器Pentium Pro，此为新的P6架构。其采用相同的0.35μm工艺与x86指令集(没有支持MMX指令集），拥有550W个晶体管，主频150MHz，处理速度达到440 MIPS（是Pentium的2~3倍性能)，插槽采用Socket8。

改进：实现CPU的L2 Cache的片内集成，采用了16KB片内L1 Cache（合计）与256KB L2 Cache。实现了3路超标量。

缺点：价格高昂。

优点：L2 Cache与CPU同速。实现了指令流水线的乱序执行。

11. Pentium II处理器：1997年，Intel发布了基于Pentium Pro的P6架构（此乃Pentum Pro的优点）的32位处理器Pentium Ⅱ，是对Pentium（P5）与Pentium Pro（P6）的改进版，其内核代号Klamath，采用0.35μm工艺并支持x86指令集、MMX指令集（此乃集合Pentium的优点，Pro中没有支持MMX），拥有750W个晶体管，主频233MHz，集成了32KB片内L1 Cache（合计）与512KB 片外L2 Cache（合计），socket为Slot1。

改进：L2缓存不再与CPU核心保持同速，采用了双重独立总线结构首次采用了最新的solt1接口标准。

缺点：L2 Cache速度较低。

优点：成本降低，其余与Pentium Pro相似。

12. Pentium III处理器：1999年，Intel发布了32位处理器Pentium Ⅲ，第一代内核代号Katmai，其实就是PentiumⅡ的简单改进版。采用0.25μm工艺，指令集为IA-32（即x86）、MMX、SSE，拥有950W个晶体管，主频500MHz，处理速度达到，socket与Pentium Ⅱ相同为Slot 1。

改进：采用0.25μm工艺，晶体管数目增加，支持SSE指令。

缺点：L2 Cache速度较低。

优点：成本降低，其余与Pentium Pro相似。

13. Pentium IV处理器：2000年，Intel发布了第一款32位Pentium Ⅳ处理器，并没有沿用Pentium Ⅲ的P6架构，而是重新设计了新微架构NetBurst。第一款Pentium Ⅳ的内核为Willametle。采用0.18μm工艺，拥有4200W个晶体管，主频1.4GHz（轻松过GHz），采用QPB型FSB，外频达到了400MHz，接口方式为Socket423。

改进：采用20级超标量pipeline，采用了奇贵的RDRAM来满足Pentium 4的带宽需求。在前端总线引入QDR技术，第一次引入SSE2指令集。

缺点：发热量大，价格昂贵，高噪音、低性能。

优点：读写达到4倍速率，浮点性能增强。

14. Pentium M处理器：2003年，Intel发布了第一代Pentium-M架构，内核代号为Banias，是对Pentium Ⅲ（即P6架构）的重大修改（Pentium Ⅲ又是在Pentium Pro基础上改的），采用130nm工艺，拥有7700W个晶体管，主频为1.3GHz~1.6GHz（性能堪比Pentium Ⅳ的2.8GHz），插槽使用了Socket 479（实际上仅478个针脚，且各针脚的定义与桌面款的Socket 478不一样）。

改进：采用12级流水线和64KB的片上L1 Cache。改进的分支预测和预取机制，

缺点：相比而言性能下降。

优点：电池续航提升，分支预测准确率比Pentium Ⅲ高20%。

15. Core i系列：

15.1 Tock——Nehalem

2008年11月，Intel推出了升级架构Nehalem，内核代号也是Nehalem，采用45nm工艺不变，是第一代Core i系列处理器。Nehalem是在Core架构上进行改进的，主频为1.06GHz~3.33GHz，核心支持2-8核，接口采用LGA1366。Intel按照i7、i5、i3的顺序发布i系列处理器，来分别对应高级、中级与入门级市场，采用相同的内核架构。

改进：引入了睿频加速技术，重新引入了超线程技术。引入三级Cache，将内存控制器集成于片内，直接取消了北桥芯片组。采用20-24级pipeline和64位宏操作融合（MOF）技术。

缺点：

优点：实现性能与省电策略的最大化，扩大了共享的L3 Cache大小，Cache效率大大提高。CPU与内存的I/O性能大大提高，总体性能大大提高。

15.2 Tock——Sandy Bridge

2011年1月，Intel推出了升级架构Sandy Bridge，内核代号也是Sandy Bridge，采用32nm工艺不变，是第二代Core i系列处理器，其架构是Nehalem的后继者。其主频为1.6GHz~3.6GHz，接口采用LGA1155，依然是64KB-256KB-2MB（per core L1, per core L2, shared L3）的Cache模式。

改进：对集成的GPU架构进行修改，采用了环形总线(Ring Bus)来链接各个core以及其L1/L2 Cache。支持了AVX 256-bit指令集，支持Intel Quick Sync Video技术。采用14到19级的pipeline。

缺点：功耗较大，会造成空载问题。

优点：核显性能增强，核显可以访问共享的L3 Cache。核之间的通信与交互变得简单。浮点性能、矩阵计算快90%。硬件视频解码加速。

15.3 Tock——Haswell

2013年6月，Intel推出了升级架构Haswell，内核代号也是Haswell，采用22nm工艺不变（更成熟，大幅改善因量子穿隧效应而导致的漏电率），是第四代Core i系列处理器。其Cache依然是64KB-256KB-2MB（per core L1, per core L2, shared L3）模式，接口采用LGA1150。

改进：主板上的电压调节模块（VRM）集成到了CPU内部，增加了两个指令集。

缺点：功耗较大，会造成空载问题。

优点：节能，核显开始模块化、可扩展的设计并使用Gen7.5架构，方便走后续暴力堆砌核显规格的道路——最高级的核显拥有40个EU，还有大容量eDRAM作为L4 Cache，可同时提升CPU与GPU性能。

15.4 Tock (Architecture)——Skylake

2015年8月，Intel推出了升级架构Skylake，内核代号也是Skylake，采用14nm工艺不变（但比Broadwell更成熟），是第六代Core i系列处理器。其最高主频高达4.5 GHz，接口采用LGA1151。Skylake同时升级了架构、工艺及核显等，四核称为标配。

改进：内存同时支持DDR3与DDR4，采用了更先进的14nm工艺，取消了Haswell中引入的FIVR电压控制模块。使用网状总线架构代替原来的ring Bus。

缺点：功耗较大，会造成空载问题。

优点：实现DDR3、DDR4 的无缝接轨，功耗降低。可超频性得到强化。

15.5 Architecture——Ice Lake

2019年8月，Intel推出了升级架构Sunny Cove，其内核代号为Ice Lake，也更新采用10nm新工艺。Ice Lake是第二代10nm工艺处理器（第一代是流产的Cannon Lake），也是第十代Core i系列处理器（另一个是同期发布的Skylake架构的Comet Lake），接口采用BGA1526。

改进：采用更深的缓冲区与更大的缓存结构，加入了AVX512指令集。支持LPDDR4X内存，集成Thunderbolt 3.0控制器。

优点：缓冲效率大大提升，功耗降低。基于机器学习来让CPU提前对未来进行动态睿频。

15.6 Optimization——Tiger Lake

2020年9月，Intel发布了代号Tiger Lake的第十一代Core系列微处理器，同时架构升级为Willow Cove，号称近年来处理器史上一次巨大飞跃：工艺到架构都有极大变化，升级力度在这几代酷睿中是非常明显的。Tiger Lake采用10nm+工艺，首次采用全新的SuperFin晶体管技术（官方表示它甚至重新定义了FinFET工艺）；得益于先进的工艺，从上代的3.9GHz一下子提升到4.8GHz，保持功耗不变的情况下提升了20%的性能。

改进：GPU使用全新的Xe架构，加入了独立的GNA 2.0（Gaussian&Neural Accelerator）单元（Intel的低功耗AI加速器），采用了大小核架构。

优点：浮点性能提升了87%，支持AI加速指令集DLBoost，针对全场景AI作了加强。能专门针对工作流负载进行优化，实现闲时低功耗待机，忙时加速AI运算。