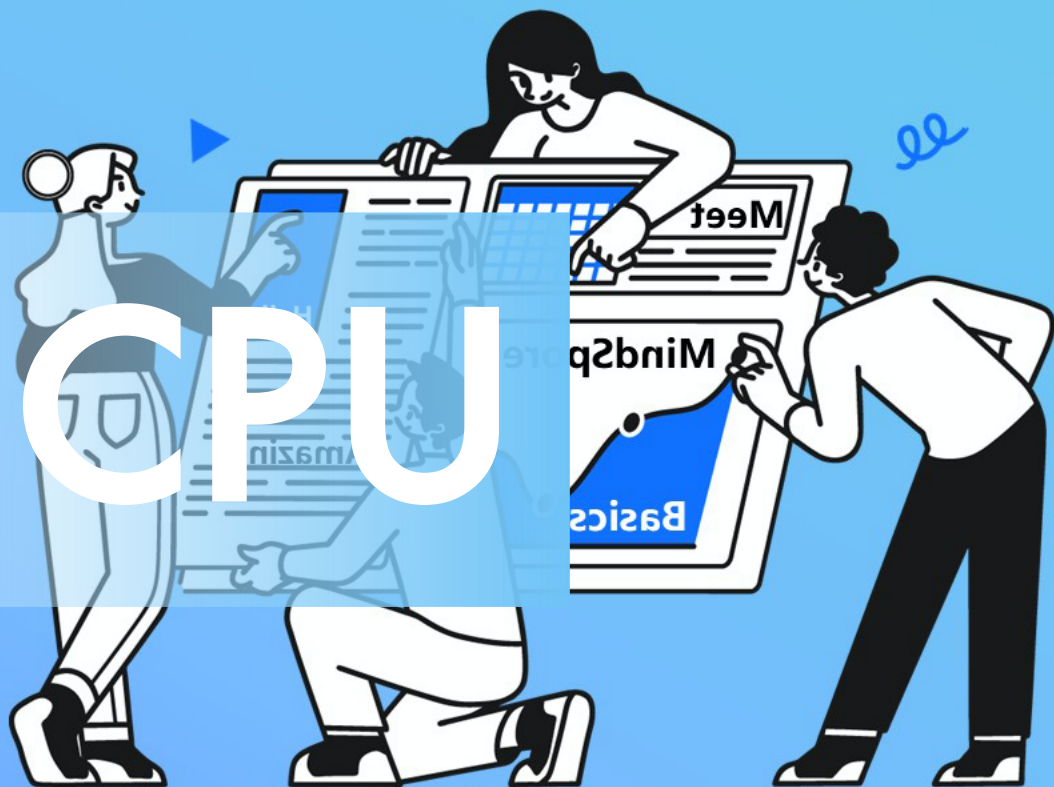


AI 芯片 – AI 芯片基础

通用处理器



ZOMI

Talk Overview

1. AI 计算体系

- 深度学习计算模式
- 计算体系与矩阵运算

2. AI 芯片基础

- 通用处理器 CPU
- 从数据看 CPU 计算
- 通用图形处理器 GPU
- AI专用处理器 NPU/TPU
- 计算体系架构的黄金10年

Talk Overview

I. 通用处理器 CPU

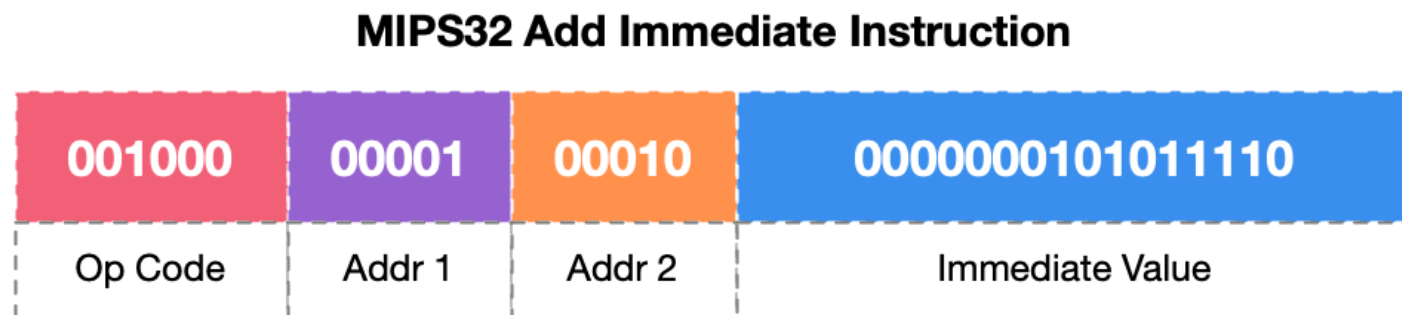
- The History – CPU 发展历史和组成
- Parallel Architecture – CPU 并行处理架构
- RoadMap of AI Chip – ISA指令集架构
- Application – CPU的应用场景

ISA

指令集架构

ISA - Instruction Set Architecture

- CPU 的全称叫中央处理器单元，通常用来区分 CPU 的标准是指令集架构（Instruction Set Architecture，简称 ISA）。
- 开发人员基于指令集架构（ISA），使用不同的处理器硬件实现方案，来设计不同性能的处理器，因此 ISA 又被视作 CPU 的灵魂。

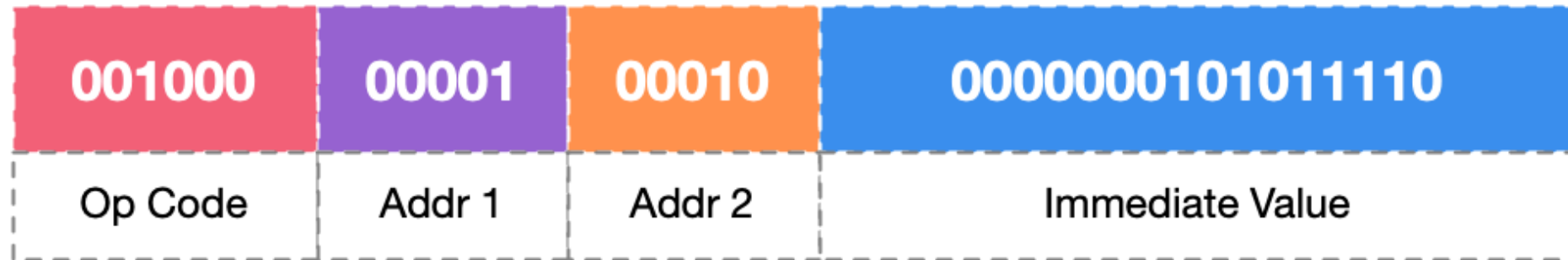


Equivalent mnemonic: **Addi \$r1 , \$r2 , 350**

One instruction may have several fields, which identify the logical operation, and may also include source and destination addresses and constant values. This is the MIPS "Add Immediate" instruction, which allows selection of source and destination registers and inclusion of a small constant.

ISA - Instruction Set Architecture

MIPS32 Add Immediate Instruction



Equivalent mnemonic: **Addi** \$r1 , \$r2 , 350

One instruction may have several fields, which identify the logical operation, and may also include source and destination addresses and constant values. This is the MIPS "Add Immediate" instruction, which allows selection of source and destination registers and inclusion of a small constant.

ISA - Instruction Set Architecture 基本分类

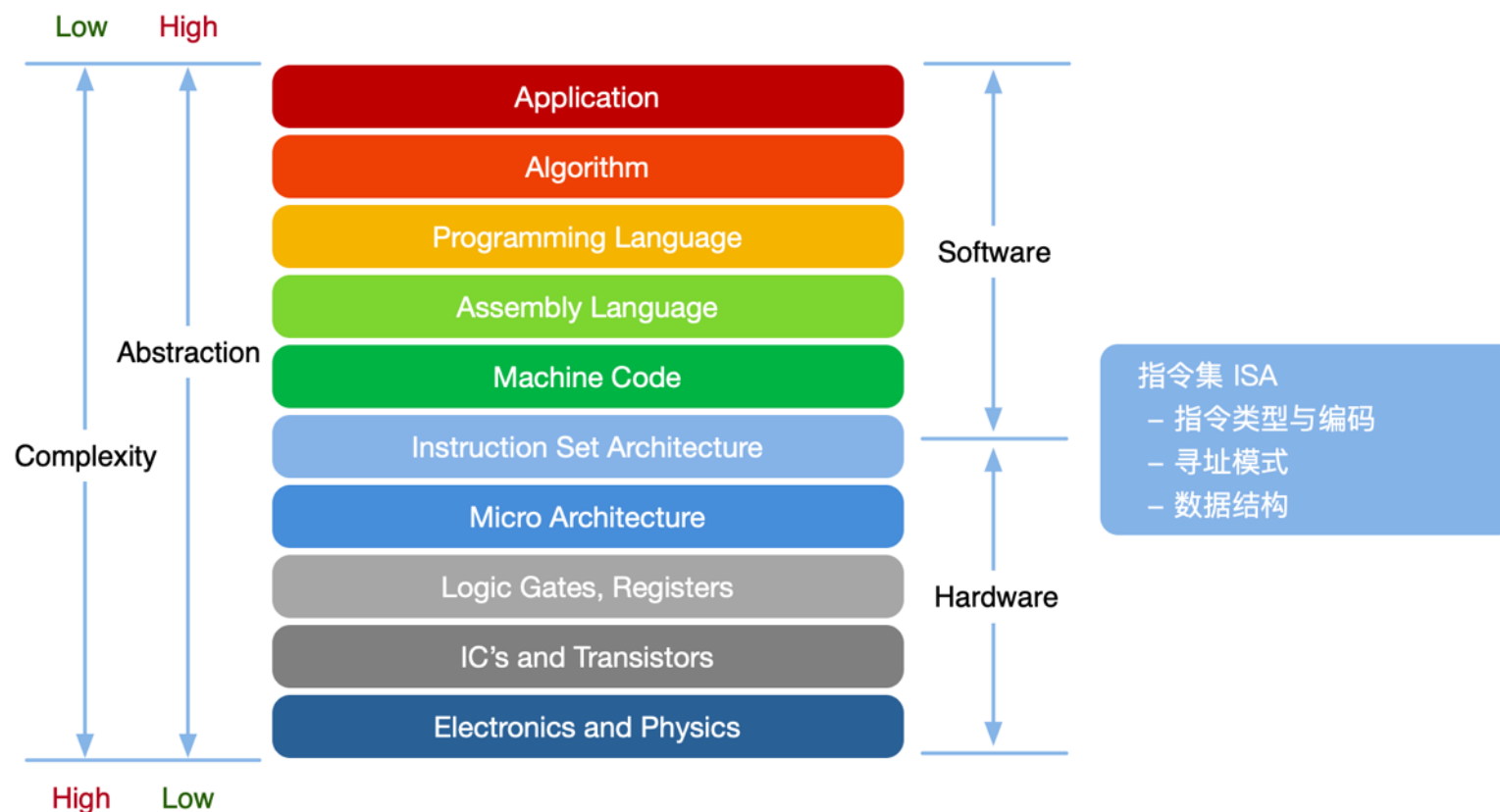
- 运算指令：在 ALU 中执行的计算操作
- 数据移动指令：读写存储操作（包括寄存器读写）
- 控制指令：更改指令执行顺序，进行程序跳转，实现 if/else，循环等

ISA - Instruction Set Architecture 声明周期

- **FETCH** : 将Memory中的指令放入Instruction Register , PC指定指令位置
- **DECODE** : 通过指令解码过程, 识别指令内容, 开启控制信号通路
- **EVALUATE ADDRESS** : 比如加载内存之前, 通过寄存器内容和偏移量获得真正的内存位置
- **FETCH OPERANDS** : 加载寄存器中的操作数, 不同指令实际执行的内容不同
- **EXECUTE** : 在ALU中执行计算逻辑
- **STORE RESULT** : 存储计算结果

ISA - Instruction Set Architecture

- 可以将指令集架构理解为一个抽象层，它是处理器底层硬件与运行在硬件上的软件之间桥梁和接口。



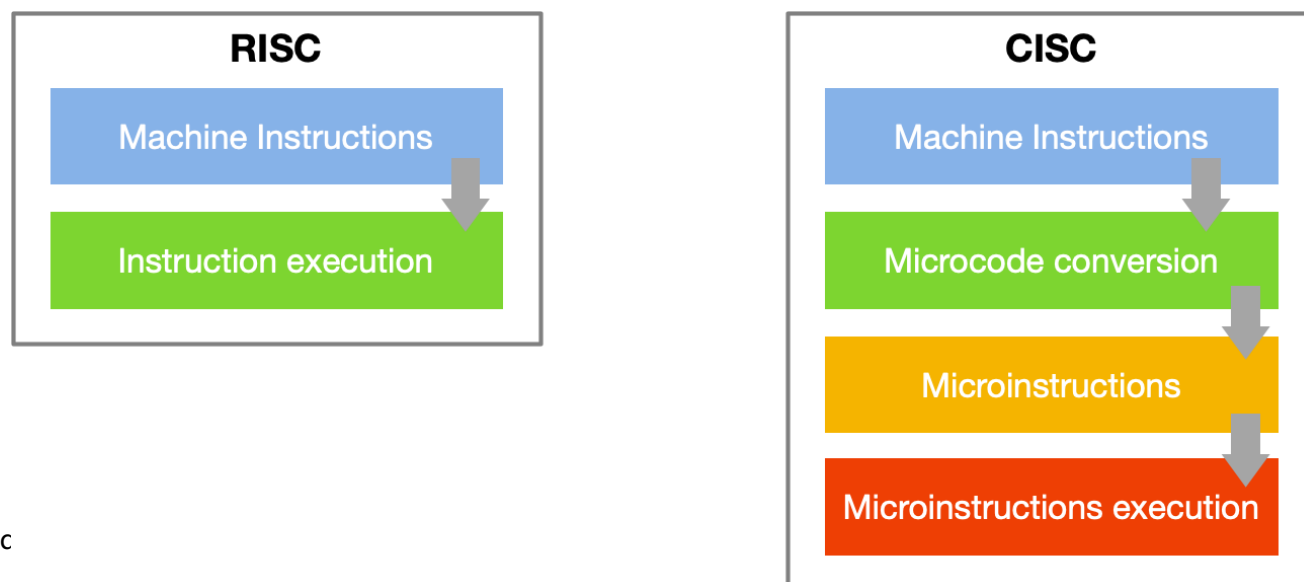
ISA 的作用

- 这个计算机架构里有多少个寄存器（ Register ）？
- 我能进行哪些运算操作？（ 有哪些指令？ADD，SUB，MUL等等）
- 如果遇到异常或者中断该怎么办？
- 数据可以有哪些类型？最多有几个字节？



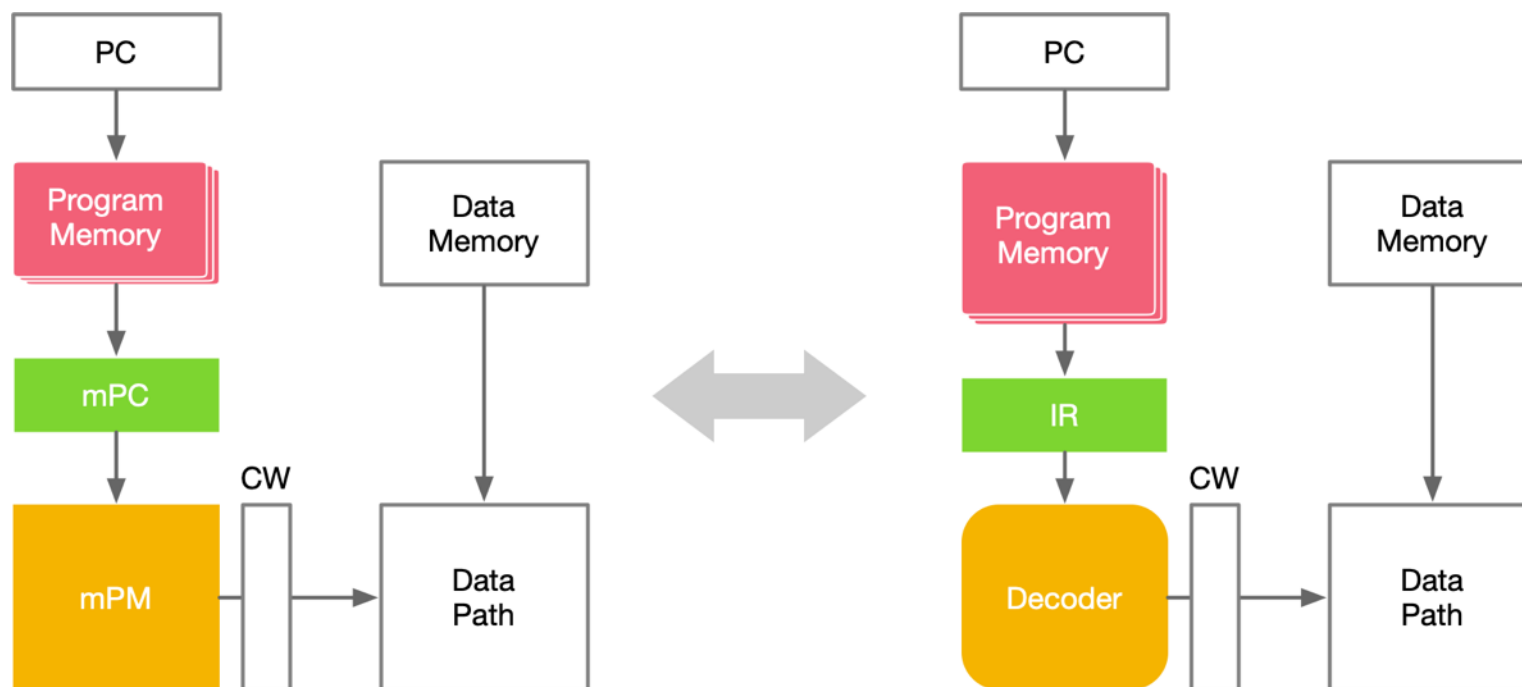
CISC 架构与 RISC 架构

- **CISC 架构**：除常用指令还包含许多不常用特殊指令。随着越来越多的特殊指令被添加到 CISC 架构中，常用程序运算指令仅占指令集 20%，80% 指令则很少用到，而这些很少用到的指令让 CPU 的设计变得极其复杂，大大增加了硬件设计的时间成本和面积开销。
- **RISC 架构**：只包含处理器常用指令，对于不常用操作，执行多条常用指令的方式来达到同样的效果。因而在 RISC 架构诞生后，移动端设备设计的 CPU 都倾向于选择使用 RISC 架构。



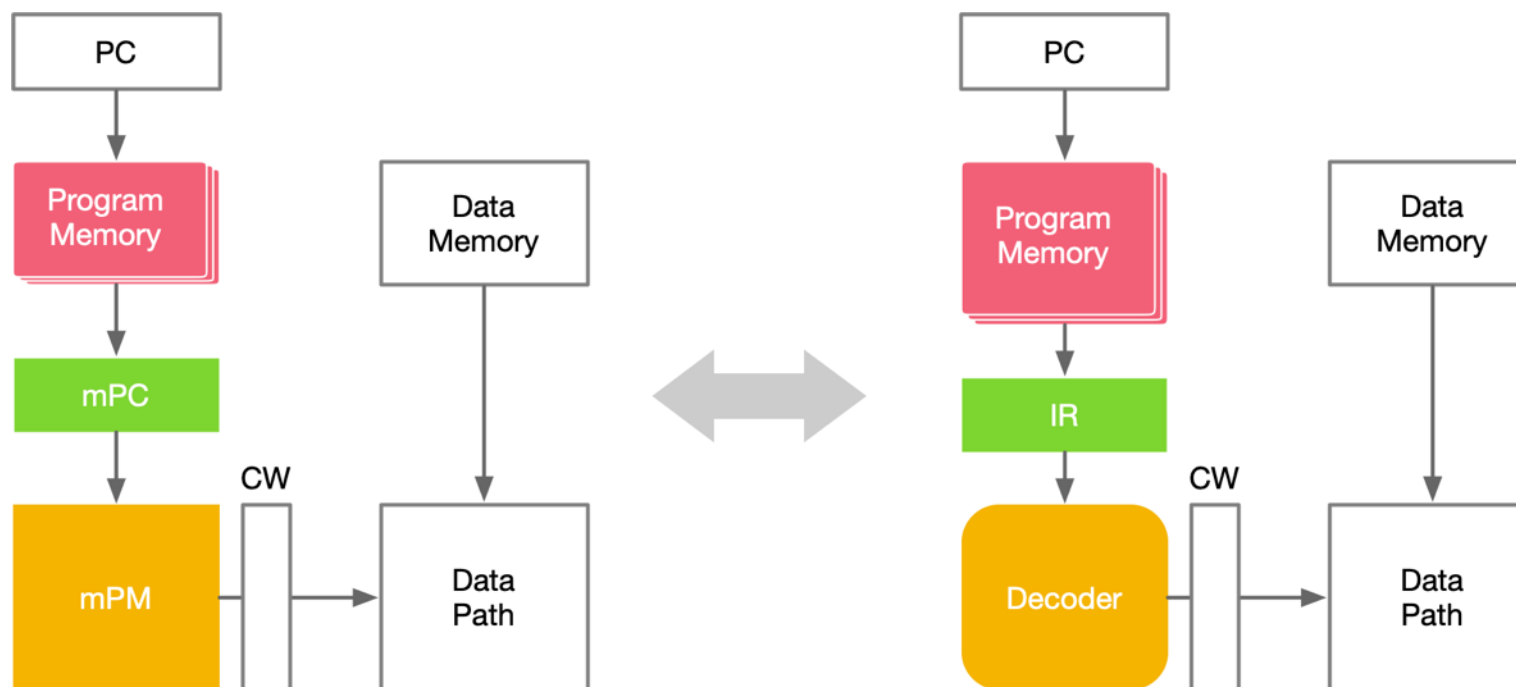
CISC 架构与 RISC 架构

- **CISC 架构**：除常用指令还包含了许多不常用的特殊指令。随着越来越多的特殊指令被添加到 CISC 架构中，常用的典型程序运算过程中用到的指令仅占指令集的20%，80%的指令则很少用到，而这些很少用到的指令让 CPU 的设计变得极其复杂，大大增加了硬件设计的时间成本和面积开销。



CISC 架构与 RISC 架构

- **RISC 架构**：只包含处理器常用的指令，对于不常用的操作，通过执行多条常用指令的方式来达到同样的效果。因而在 RISC 架构诞生后，移动端设备设计的 CPU 都倾向于选择使用 RISC 架构。



CISC vs RISC

	CISC	RISC
指令系统	复杂，庞大	简单，精简
指令数据	>200	<100
指令长度	不固定	定长
可访存指令	不加限制	只有Load/Store指令
指令执行时间	相差较大	大部分在一个周期内完成
指令使用频率	相差较大	都比较常用
通用寄存器数	较少	多
目标代码	难以用优化编译生成高效的目标代码程序	采用优化的编译程序，生成代码较为高效
控制方式	微程序控制	组合逻辑控制
指令流水	可以通过一定方式实现	必须实现

ISA 种类

- CPU 于上世纪 60 年代问世，已发展几十年，有几十种不同的指令集架构相继诞生或消亡。

指令集架构	指令集架构	公司
X86	CISC 架构个人计算机的标准处理器架构	Intel/AMD
ARM	32 位和 64 位 RISC 系列声名显赫，无处不在	ARM
RISC-V	完全开放的指令集，源自名校，兴于开源	RISC-V 基金会
SPARC	高性能 RISC 架构的代表针对服务器领域设计	Sun
Power	RISC 架构高性能领域优势明显应用于高端服务器	IBM
ARC	32 位 RISC 架构以极高的能效比见长	Synopsys
MIPS	简洁优化 RISC 架构，广泛用于嵌入式设备及消费领域仅次于 ARM	/
Alpha	64 位 RISC 架构处理器多应用于企业级服务器但价格高昂、部署困难，淡出市场	/

CPU Application

应用场景



应用场景

CPU 典型应用领域及主流架构

服务器领域

主流架构：

- Intel X86 架构高性能CPU占据垄断地位



PC领域

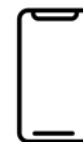
主流架构：

- Intel / AMD x86 高性能CPU占据垄断地位



嵌入式领域

- 移动设备：ARM Cortex-A架构
- 嵌入式实时设备：ARM / RISC
- 泛嵌入式设备：ARM Cortex-M / RISC

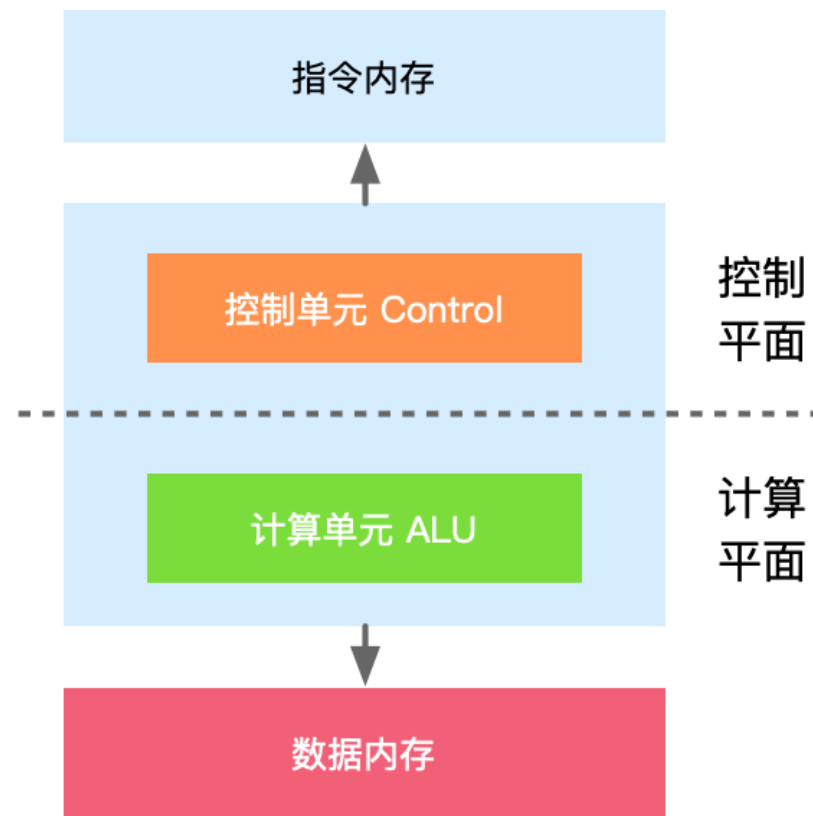


Summary

总结

总结

- CPU 具备图灵完备，可自运行的处理器。其主动从指令内存读取指令流，然后译码后执行；指令执行会涉及到数据的载入(Load)、计算和存储(Store)。
- ISA 是计算机体系结构与编程相关的部分，定义了：指令集、数据类型、寄存器、寻址模式、内存管理、I/O模型等。
- 可以把 CPU 简单地分为控制平面和计算平面两部分，CPU 作为指令流驱动计算的处理引擎。



引用

1. [Processor Architecture Design Practices: survey & Issues](#)



BUILDING A BETTER CONNECTED WORLD

Ascend & MindSpore

20

www.hiascend.com
www.mindspore.cn



BUILDING A BETTER CONNECTED WORLD

THANK YOU

Copyright©2014 Huawei Technologies Co., Ltd. All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.