

计算机体系结构基础

（期末复习）

胡伟武、苏孟豪

本课程的目的

- 建立系统观
- 掌握一些概念
- 掌握一些能力

什么是计算机体系结构

- 计算机体系结构（**ComputerArchitecture**）是描述计算机各组成部分及其相互关系的一组规则、方法和机制，是程序员所看到的计算机的属性，即概念性结构与功能特性。
 - 计算机体系结构主要研究内容包括指令系统结构（**Instruction Set Architecture**）和计算机组织，微体系结构（**Microarchitecture**）是处理器级别的计算机组织。
 - 高性能、低功耗、低成本是计算机体系结构的主要设计目标。
 - 冯诺依曼结构的存储程序和指令顺序执行是现代计算机体系结构的基础。

计算机体系结构的表現方式

- 计算机体系结构可以有不同层次和形式的表現方式。
 - 计算机体系结构通常用结构框图来表示，结构框图中的方块表示计算机的组成部分（如运算器、寄存器、存储器、输入/输出设备等），线条和箭头表示指令和数据在功能模块中的流动，结构框图可以不断分解一直到门级或晶体管级。
 - 计算机体系结构也可以用高级语言如C语言来表示，形成体系结构模拟器，用于对体系结构进行性能评估和分析。
 - 以Verilog为代表的硬件描述语言描述的体系结构可以通过EDA仿真工具进行功能验证和性能分析，并可以通过EDA综合布局布线工具转换成门级以及晶体管级网表并最终转换成版图，用于芯片制造。

本课程主要内容

- 第一部分：引言（第1章）
 - 体系结构研究内容、主要性能指标、发展趋势以及设计原则
- 第二部分：作为软硬件界面的指令系统结构（第2-4章）
 - 建立高级语言（如C语言）与指令系统结构的关系
 - 指令反映了结构设计者对应用的深刻理解，不仅仅是操作编码
 - 指令系统结构、特权态指令系统结构、软硬件协同
- 第三部分：CPU、GPU、南北桥协同的硬件结构（5-7章）
 - 冯诺依曼结构的具体体现；搞清楚计算机内部包括CPU、GPU、内存、IO之间是如何协同完成软件规定的各种操作的
 - 计算机硬件结构、计算机接口及总线、计算机系统上电启动过程

本课程主要内容（续）

- **第四部分：CPU的微结构（第8-9章）**
 - 建立指令系统和晶体管之间的“桥梁”，硕士课程的重点内容
 - 运算器设计、指令流水线（控制器设计）
- **第五部分：并行处理结构（第10-11章）**
 - 现代计算机通过多层次的并行性开发来提高性能
 - 应用程序的并行行为、多核处理器
- **第六部分：计算机性能分析（第12章）**
 - 性能不由一两个具体指标决定，而是若干因素综合平衡的结果
 - 计算机性能指标（计算机性能基准测试程序、计算机性能比较）、计算机性能分析方法

第01章：引言

- 计算机体系结构的研究内容
 - 冯诺依曼结构存储程序和指令驱动执行是计算机体系结构的基础
 - 计算机系统的四个层次及三个界面：应用、OS、硬件、工艺
- 衡量计算机的指标
 - 衡量性能的指标：性能最本质的定义、IPC、影响IPC的因素
 - 功耗的组成：静态功耗、动态功耗
- 计算机体系结构的设计原则
 - 平衡性、局部性、并行性、虚拟化
- 计算机体系结构的发展

第02章 指令系统

- 指令系统简介
 - 软硬件界面，结构设计者对应用的深入理解，指令操作和运行时环境
- 指令系统的设计原则
 - 兼容性、通用性、高效性、安全性
- 指令系统的演变
 - 指令集分类：CISC、RISC、VLIW
 - 存储管理：连续实地址、段式、页式
 - 运行级别：用户态、核心态
- 指令系统组成
 - 地址空间：寄存器空间、存储空间，堆栈型、累加器型、寄存器型
 - 操作数：操作数的存储（寻址方式、大小尾端），操作数的特征
 - 指令操作和编码：运算指令、访存指令、转移指令、系统管理指令
- RISC指令系统比较：“亲兄弟”和“表兄弟”
- C语言的机器表示
 - 如C语言的不同变量映射到地址空间的不同段

第03章 特权指令系统

- 特权指令系统简介
- 异常与中断
 - 异常分类：外部事件、指令执行中的错误、数据完整性问题、地址转换异常、系统调用和陷入、需要软件修正的运算
 - 异常处理：异常处理准备、确定异常来源、保存执行状态、执行异常处理、恢复执行状态、返回正常执行流
 - 中断：可屏蔽和不可屏蔽中断、中断优先级
 - 中断传递机制：中断线和MSI
- 存储管理
 - 虚拟存储原理：存储管理作用、页式存储管理，地址转换和TLB
 - MIPS处理器对虚存系统的支持
 - LINUX操作系统的存储管理：
 - `Array=(int*)malloc(0x1000); For (I=0;I<1024;I++) Array[i] = 0;`
 - TLB的性能分析和优化

第04章 软硬件协同

- 应用程序二进制接口
 - ABI规范的内容、MIPS寄存器约定
- 中断的生命周期
 - 中断的上、下半部
- 系统调用过程
- 同步与通讯
 - 用LL/SC指令实现锁机制

第05章：计算机组成原理和结构

- 计算机组成原理和结构
 - 冯诺依曼结构的五大组成部件
- 计算机硬件结构的演进
 - 四片结构、三片结构、两片结构、单片结构
- 处理器和IO间通信
 - IO寄存器寻址方式：专用指令、地址空间映射
 - CPU与IO间的同步：查询方式、中断方式
 - CPU与IO间的通信：PIO方式、DMA方式
- 计算机系统主要组成部件
 - DRAM的单元结构和读写过程

第06章 计算机总线接口技术

- 总线概述
 - 总线规范：机械层、电气层、协议层、架构层
- 总线分类
 - 片上总线/系统总线/内存总线/I0总线
 - 并行/串行，共享/点对点，全局时钟/源同步
- 总线举例
 - 片上总线：AXI
 - 内存总线：DDR3
 - 系统总线：HT
 - I0总线：PCIE

第07章：计算机系统启动过程分析

- 处理器核初始化
 - 第一条指令从哪里取的？什么时候把内核拷贝到内存？
- 总线接口初始化
 - 配置空间，IO空间，Memory空间
- 设备探测及驱动加载
 - PCI设备的探测
- 多核启动过程

第08章：运算器设计

- 二进制与逻辑电路
 - 计算机中数的表示：原码/补码/移码、IEEE754浮点数表示
 - MOS晶体管工作原理：NMOS管和PMOS管
 - CMOS逻辑电路：组合逻辑和时序逻辑、CMOS晶体管级电路图
- 简单运算器设计
 - 定点补码加法器：先行进位
 - 减法运算实现：减数按位取反，进位加一
 - 比较运算实现、移位器
- 定点补码乘法器
 - 补码乘法器
 - Booth乘法器：两位一乘算法
 - 华莱士树：进位传递过程

第09章：指令流水线

- 简单的处理器数据通路
- 流水线处理器
 - 五级流水线
- 解决指令相关
 - 数据相关(WAW、WAR、RAW)，控制相关、结构相关
 - 时空图
- 指令调度技术
 - 软件调度（循环展开），硬件调度
- 提高流水线效率的技术
 - 多发射、乱序执行；
 - 转移猜测、存储层次

第10章：并行编程基础

- 程序的并行行为
 - 指令并行，数据并行，任务并行（多线程、多进程）
- 并行编程模型
 - 单任务数据并行模型
 - 多任务共享存储编程模型
 - 多任务消息传递编程模型
- 典型的并行编程环境
 - SIMD编程
 - Posix编程环境
 - OpenMP编程环境
 - MPI编程环境

第11章：多核处理结构

- 多核处理器的发展演化
 - 分类：通用/专用、同构/异构、多核/众核
- 多核处理器的访存结构
 - 片上Cache结构：私有、片上共享、片间共享
 - 存储一致性模型：顺序/处理器/弱/释放一致性
 - Cache一致性协议：目录/侦听，write-invalidate/write-update
- 多核处理器的互连结构
 - 片上总线、交叉开关、片上网络，路由算法，流量控制
- 多核处理器的同步机制
 - 锁同步、栅障同步、事务内存，“读-改-写”指令、“LL/SC”指令
- 典型多核处理器

第12章：计算机性能分析

- 计算机系统性能评价指标
 - 不同系统关注的性能，影响性能的因素，性能最本质的定义
 - 名词解释：CPI、MIPS、MFLOPS、TPS、FPS、MBPS、MHz
 - 并行系统的评价指标：加速比，amdahl定律
- 性能测试程序集
 - 常见的基准测试程序：Coremark, SPEC CPU, EEMBC, Linpack
- 计算机性能分析方法
 - **分析**建模的方法：概率模型、队列模型、马科夫模型、Petri网模型
 - **模拟**建模的方法：系统vs.部件，全系统vs.用户态，执行驱动vs.踪迹驱动，时钟驱动vs.事件驱动，功能模拟器vs.时序模拟器
 - 性能**测量**的方法：硬件监测、软件监测、微码插桩
- 性能分析和测试实例

需要掌握的重点能力（一）

- 掌握从键盘操作到显示变化的整体软硬件过程
 - 如：PPT翻页过程
- 掌握基于MIPS处理器的Linux操作系统TLB例外过程
 - 指出一段程序共产生多少次TLB refill和TLB invalid异常（可能包含TLB替换和进程切换）
- C语言与指令系统的关系
 - C语言的不同变量（全局、局部等）映射到地址空间的哪些段？
 - 函数调用、系统调用和中断处理都需要上下文切换，请结合MIPS O32的ABI说明上述三种上下文切换时保留现场有什么不同（内容、位置）？

需要掌握的重点能力（二）

- 系统初始化时PCI设备的探测
 - 如系统中有关IO设备是如何探测到的
- IO通信时DMA传输过程
 - 如简单描述从硬盘读一个文件的DMA过程
- 内存总线接口
 - 给定一段DRAM总线波形图，给出DRAM访问的读写地址（要求熟悉DRAM总线各信号）
- AXI总线接口
 - 用Verilog写出AXI接口的RAM（给出AXI关键信号）

需要掌握的重点能力（三）

- 计算机中数据格式
 - 定点补码、原码格式
 - 浮点格式
- 读懂CMOS电路
 - 给出一个晶体管电路，给出真值表和等价的逻辑表达式
- 先行进位加法器
 - 用Verilog写出指定位数的先行进位加法器，或画指定位数的加法器的结构，并给出最长延迟（门的级数）
- 定点补码乘法器：Booth编码和华莱士树
 - 给出指定个数相加的华莱士树的最小层数、利用半加器能增加多少进位（不改动全加器和改动全加器）

需要掌握的重点能力（四）

- 指令流水线
 - 给出一段程序，要求画出五级静态流水线时空图（可要求指出阻塞原因），引入前递、提前计算跳转地址等机制后的流水线时空图
- 转移猜测
 - 结合转移猜测算法，计算一段小程序的命中率
- Cache结构
 - 访存地址中Tag、index、offset与Cache结构的对应关系

需要掌握的重点能力（五）

- **OpenMP、MPI的基本思想，Amdahl定律**
 - 结合Amdahl定律，计算一段OpenMP程序的加速比
- **原子指令、同步操作、自旋锁**
 - 利用LL/SC指令实现锁
- **性能比较分析**
 - 计算不同计算机的MIPS和IPC，并比较不同计算机的性能
- **基准程序**
 - 分析EEMBC和SPEC CPU2000行为特性上有什么差别？

谢 谢