

体系结构 — 考试重点整理

考试格式与复习范围

- **分值:** 单选 30, 简答 20, 问答题 (4题) 50。
- **复习范围:** 四个大题与作业题2 3 5 8题型相似。

流水线 (Pipelining)

- **考点:**
 - 理解流水线的基本思想：提高吞吐量而非单条指令延迟。
 - 为什么时钟周期不能简单等于单周期时钟/N：寄存器开销、最长级段决定周期、旁路/前递等额外线路。
 - 不考具体电路图；主要为常识性理解（如为什么 CPI>1 以及流水线深度的利弊）。
- **需要掌握:** 概念、流水线深度带来的 overhead 与收益递减。

指令相关性以及危害

数据相关性？控制相关性？结构相关？概念
危害？概念

数据相关三种：wAR WAW RAW（真相关）

只有RAW会让流水线停顿，WAR WAW是因为寄存器数量不够产生的可以通过重命名解决。

存储层次 (Cache & 虚拟内存)

- **Cache (一个大题)**
 - **如何工作与设计:** 采取多层次，能讲清工作流程（访问 -> tag 比较 -> hit/miss -> 替换/写回策略）。
 - **局部性:** 时间局部性？空间局部性？概念。在哪儿表现？（得对着代码看得出例子）如何利用时间局部性（解：最近访问数据放到cache这种访问速度快的地方）/空间局部性（解：将与近期访问数据相邻的数据放到cache这种快的地方）？
Cache相关概念 什么是block? Cache怎么寻址？(index bits) 直接映射？组相联？全相联？
 - **替换策略:** 只考 FIFO 与 LRU（理解区别与何时使用）。
 - **Cache hit:** 掌握基本含义与结果。
 - **Write miss 处理:** 基本掌握处理write miss?(write-allocate / no-write-allocate) 与处理 Cache hit(write-back / write-through)的基本效果。
 - **增大 cache size 的影响:** 知道随着 block size/ cache size / 关联度变化的典型曲线与趋势（大致理解，不需精确图表）。
 - **如何提升cache性能** 降低 miss rate 降低 miss penalty 降低 hit latency 增大 cache size？增加关联度？增加block size？三者各自对三种miss (compulsory capacity conflict) 的好处（只知道比如+block size能）
 - **块大小:** 通常能从题目描述/序列辨识出合理的块大小选择。
- **虚拟内存 (一个大题) — 重点掌握**
 - **概念与好处:** 为什么要有虚拟内存；物理地址 (PA) 与虚拟地址 (VA) 的含义及优点（隔离、内存扩展、共享等）。

- 地址转换：知道 VA 到 PA 的转换过程及涉及的页表。
- 页表结构：理解单级与多级页表（多级页表要“看一眼”懂其思想）。
- TLB（重要）：TLB 缓存了什么（页表项）；何时访问 TLB 或访问页表；TLB 替换策略仅考 LRU 与 FIFO；TLB miss 的基本处理流程与软/硬件方案的优缺点。
- 注意：不涉及别名（Aliasing）细节。

乱序执行（Out-of-Order）与相关结构

- 考点：

- Tomasulo 算法的核心思想（保留站、发射、执行、写回的基本流程）。
- 寄存器重命名与乱序发送的概念。
- 精确异常的要求与实现思路。
- Reorder Buffer（ROB）的概念、作用，与保留站（Reservation Station）的区别。
- 考法：可能出现小题考察概念，不会是大题。

分支预测

- 考点：

- 2-bit 饱和计数器要重点掌握（作业题相关）。
- 其他复杂预测器不涉猎或简略。
- 分支预测题目与作业题相似，复习作业非常重要。

多发射 / 超标量 / VLIW

- 理解概念：超标量、超长指令字（VLIW）、多发射的基本含义与区别。

多线程（细粒度/粗粒度）

- 掌握：概念与优缺点，何时使用各自策略。

Cache 一致性与内存一致性

- Cache 一致性（重点）：

- 理解基于 invalidate 的协议与基于 update 的区别（重点看 invalidate 机制）。
- 必考：MSI 协议（要看转换状态的逻辑与原因），MESI 也要了解；不考 MOESI 等更复杂变体。
- 明白为什么从 MSI 扩展到 MESI（性能/优化原因）。

- 内存一致性（Memory Consistency）：

- 与 Cache 一致性不同点要能区分。
- 顺序一致性（Sequential Consistency）要知道定义；弱一致性细节可选不深入。

GPU（重要）

- 设计思路三条要点（在复习时重点关注架构思路）：1.简化流水线，增加核心数；2.单指令多线程 SIMT；3.同时驻留大量线程
- 大量并行线程与层次化硬件支持（线程块/线程/warp 等概念）。

- 硬件对控制流分歧 (branch divergence) 的处理思路 (了解如何影响性能) 。
- 内存访问模式优化与带宽/延迟权衡。
- **题型:** GPU 可能有作业题, 控制流分歧与线程行为相关题目要看一眼。

其他要点与复习建议

- **PPT / 作业:** 四个大题通常就是四个作业题, 优先复习作业答案与讲义中的例题。
 - **重点归纳:**
 - 虚拟内存大题优先复习 (页表、TLB、地址转换) 。
 - Cache (工作原理、替换策略、写策略) 为另一个大题。
 - 分支预测 (2-bit 饱和计数器) 与 GPU/乱序相关题常出现在作业中。
 - **答题技巧:** 能讲清概念、画出数据流或简要步骤图 (如地址转换流程、TLB 访问顺序、Cache 访问路径) 通常足够。
-