

并行处理 2024 秋季学期复习资料

1. 从上层应用出发的并行政程序的两种通用模型是什么？请列出并分别解释这两种模型。
2. 列出现代处理器并行执行的主要形式，并分别解释。
3. 分析多线程的收益和代价，并举例吞吐导向的多线程代表架构。
4. Flynn 分类法是如何对并行分类的？
5. 多核有哪几种通信方式？
6. 列举减少访存延迟和隐藏访存延迟的方法。
7. SMP 和 NUMA 架构分别是什么，各自有什么优缺点。
8. 并行编程模型分别是哪三种，并分别阐述三种模型各自的特点。
9. 请举一个常用的混合编程模型的例子？其好处是什么？
10. 基于共享地址空间的通用同步原语有哪些？具体进行解释。
11. 在讨论局部性时会从哪两个维度上进行讨论，两种局部性分别指的是什么，cache 利用了哪种局部性原理？
12. 导致 cache miss 的原因是什么？并分析如何避免/减少每种 cache miss？
13. 请列举出降低通信开销的几种方法？

14. 并行编程中，造成竞争的原因有哪些，如何减小竞争？
15. 介绍为了使工作负载更均衡，在任务调度时可以采取的机制，以及可能的困难。
16. 导致应用可扩展性差的因素有哪些？并从中选出两个因素分析如何发现这种因素以及如何解决？
17. 请解释强可扩展性和弱可扩展性。
18. 导致并行计算无法达到理想加速比的并行性能开销有哪些？
19. 请画出 VGG 和 MobileNet 两个神经网络模型推理对应的 roofline 模型，并标出 VGG 和 MobileNet 在图中的位置。
20. 请结合 roofline 模型，在下述配置的处理器中优化 Stencil 程序。
21. 请简述并程序 Benchmark 的选取原则是什么。
22. SIMD 和 SIMT 架构分别是什么，请分别解释。
23. 简述线程分组的两种方式，并分析动态线程分组的缺点。
24. 介绍一下 CUDA 编程中，并发线程的层次结构，以及 CUDA 对并行提供的支撑。
25. 请列举至少三种提高 CUDA 代码效率的方法。
26. 列举四种常见的共享存储多处理器架构，并分别解释。

27. 解释存储一致性模型、顺序一致性模型。
28. 在顺序同一性下, (A,B)的可能结果包括哪几种? 并举例说明对应的实际执行顺序(每种结果举 1 个例子即可)。不可能的结果是什么? 如果实际执行的顺序是 $1b \rightarrow 1a \rightarrow 2b \rightarrow 2a$, 满足顺序同一性吗? 解释原因。
29. 举例几种常见的互联拓扑结构, 并给出其网络直径和和对分带宽分析。
30. 论述互连网络中, 死锁出现的原因和避免的方法, 给出一种避免方法的示例。
31. 片上网络的流控按粒度划分可以分为哪几类, 并分别介绍一种对应的流控技术
32. 分析 N 维、每维上有 k 个节点的 Torus 片上互连拓扑结构的对分带宽、饱和吞吐、网络直径和平均跳步。
33. 列举常用的利用空间局部性的方法。
34. 列举事务性内存的优势。