

高等计算机体系结构，2021 年春季

作业 6：预取和并行

主讲教师：栾钟治

助讲教师：杨海龙；助教：孙庆骁

作业下发时间：2021 年 5 月 28 日

作业回收时间：2021 年 6 月 11 日

1 预取 I 20 分

假如你是一位架构师，正在为你的机器设计预取引擎。你先在机器上使用跨度预取器执行了 A 和 B 两个应用。

应用 A:

```
uint8_t a[1000];
sum = 0;
for (i = 0; i < 1000; i += 4)
{
    sum += a[i];
}
```

应用 B:

```
uint8_t a[1000];
sum = 0;
for (i = 1; i < 1000; i *= 4)
{
    sum += a[i];
}
```

i 和 sum 在寄存器中，数组 a 在内存中，一个 cache 块大小为 4 个字节。

(a) 使用跨度预取器，应用 A 和 B 的预取精度和覆盖率分别是多少？这个跨度预取器检测两次连续访存的跨度，从当前访问的 cache 块按照这个跨度预取下一个 cache 块。

(b) 请分别为应用 A 和 B 建议能获得更好的精度和覆盖率的预取器

i) 应用 A

ii) 应用 B

2 预取 II 25 分

你跟你的同学一起设计一个预取器，这台机器使用单核、L1 和 L2 cache 以及 DRAM 内存。我们需要分析不同的预取器和可能的 tradeoff。

在本题中，我们要计算预取器在达到稳定状态后的预取精度、覆盖率和带宽开销，所以，**所有计算都不包括最开始的 6 次请求，这 6 次请求作为预取器的训练集。**

(a) 你首先设计一个跨度预取器，观察最后三次 cache 块请求，如果最后三次请求的跨度是常数，预取器将会使用这一跨度预取下一个 cache 块。

你执行了一个应用，它具有如下的访存模式 (这些是 cache 块地址):

A, A+1, A+2, A+7, A+8, A+9, A+14, A+15, A+16, A+21, A+22, A+23, A+28, A+29, A+30...

假设这个模式持续了很长时间。

计算你的跨度预取器对于这个应用的精度和覆盖率

(b) 你的同学设计了一个新的预取器，当有一个 cache 块访问时，预取紧接着的 N 个 cache 块

(i) 如果用他的预取器执行你刚刚执行过的应用，预取覆盖率和精度分别是 66.67% 和 50%，N 是多少？

(ii) 假如我们将带宽开销定义为：有预取器时所有 cache 块的请求数/没有预取器时所有 cache 块的请求数，那么你同学的预取器在执行刚才那个应用时的带宽开销是多少？

(c) 你的同学希望改进他的预取器对于刚才那个应用的覆盖率，他可以容忍带宽开销最多两倍。请问他能做到吗？为什么可以/不可以？

(d) 对于上面的应用，如果想获得 100% 的覆盖率，N 最小得是多少？这个时候的带宽开销是多少？

3 Cache 一致性 10 分

(a) MESI cache 一致性协议比 MSI 协议好在哪里？

(b) 你想要利用 MESI 置无效协议设计一个基于目录的 cache 一致性系统，在特定的工作负载下，系统表现得很糟糕，经过仔细的分析，你发现有 4 个节点持续的发出对某个 cache 块的置无效请求，这什么情况？

(c) 如何解决这一问题？

4 一致性协议 15 分

假设有一个多处理器系统，系统有 512 个处理器，每个处理器有 1MB 的私有写回方式的 cache，每个 cache 块 64 字节，主存大小为 1GB。

(a) 如果我们基于 MESI cache 一致性协议设计了监听总线，需要多少状态位才能够实现这个一致性协议？这些状态位放在哪？

(b) 如果用基于目录的 cache 一致性协议替换，需要多少状态位？这些状态位在哪？

(c) 对于这个系统，你会选择哪一个协议？为什么？

5 并行加速比 30 分

假如你是一家公司的程序猿，你被要求并行化一个老程序以使它能够在现代多核处理器上跑得更快。

(a) 你并行化了这个程序，然后发现它对于单线程版本的加速比相比于处理器个数的增加而言相差很多。你发现在每个核的数据 cache 中有大量的 cache 无效存在，什么样的程序行为导致了这种现象？(请用 10 个字左右简要说明)

(b) 你修改了程序以解决这个性能问题，然后你发现程序在每个并行计算之后的一个单个线程都会更新一个全局状态，因此导致性能的下降。你的程序有 90% 的工作是并行的(按照处理器个数 X 秒计算得出)，另有 10% 的工作是串行的，并行部分是完美的并行。如果多核处理器核数无限，程序的最大加速比可以到多少？

(c) 如果要获得 4 倍的加速比，应该有多少处理器？

(d) 为了使你改写的程序更高效，公司决定设计一款专门的异构处理器。这款处理器由一个大核(执行代码更快，但是占据的片上面积更大)和多个小核(执行代码更慢，但是消耗面积更小)共享处理器的片上空间。

你的程序并行部分的所有线程将只会在小核上执行；程序的串行部分将有一个线程执行在大核上。核的性能(执行速度)与它的面积的平方根成正比。

假设芯片面积有 16 个单元可用，一个小核至少占用 1 个单元，大核可以占用任意数量的单元。同时假设没有被大核使用的面积会被小核填满。

(i) 如果想让你的程序获得可能的最快执行速度，大核需要多大？

(ii) 如果所有 16 个单元全部拿来用做小核，这个处理器就变成了同构的多核处理器，对于你的程序而言，它的加速比是多少？假设串行部分跑在一个小核上，并行部分跑在所有 16 个小核上。

(iii) 在串行部分是 10% 的情况下，使用异构多核(大小核)处理器是有意义的吗？为什么是/不是？

(e) 现在你继续优化了你的程序，使得串行部分仅占 4%(剩下 96% 是并行部分)。

(i) 这个时候大核应该有多大(占多少单元)？

(ii) 大核这么大的时候加速比是多少？

(iii) 假如此时我们采用 16 个小核的同构多核处理器，你的程序的加速比是多少(假设串行部分跑在一个小核上，并行部分跑在所有 16 个小核上)?

(iv) 在串行部分是 4%的情况下，使用异构多核(大小核)处理器还是有意义的吗? 为什么是/不是?