Multi Banked Caches For Increasing Cache Bandwidth

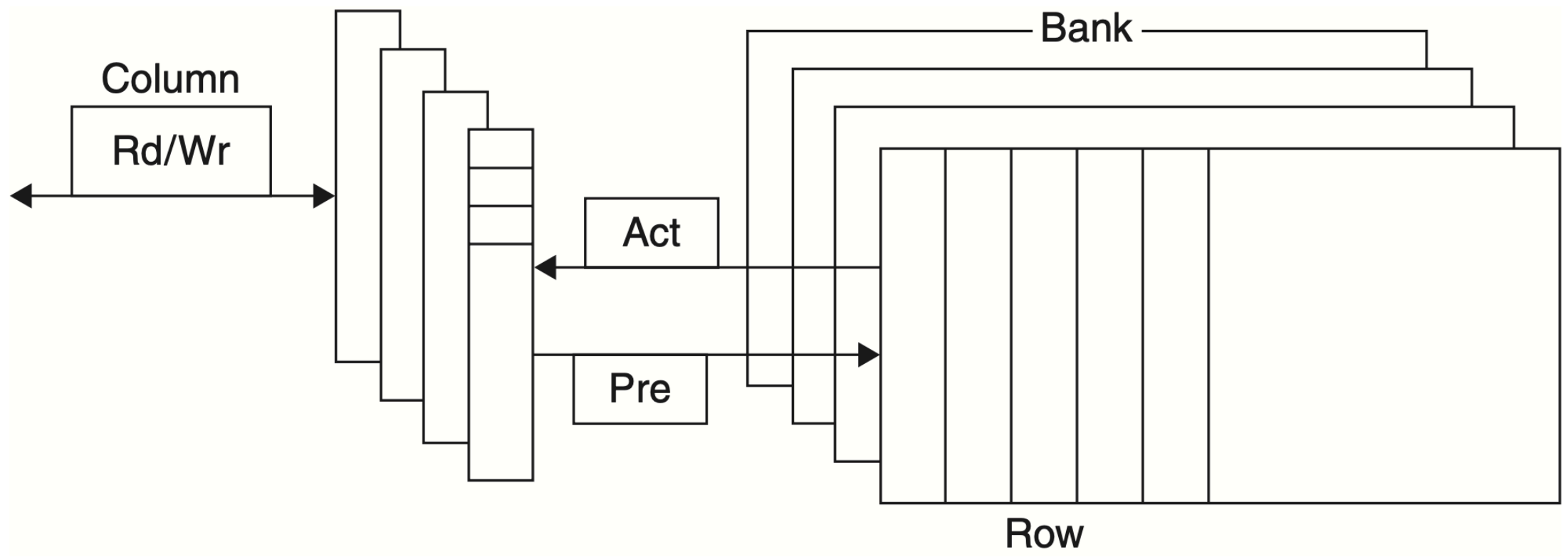
What is Multiple Banks?

- **Multiple Banks**

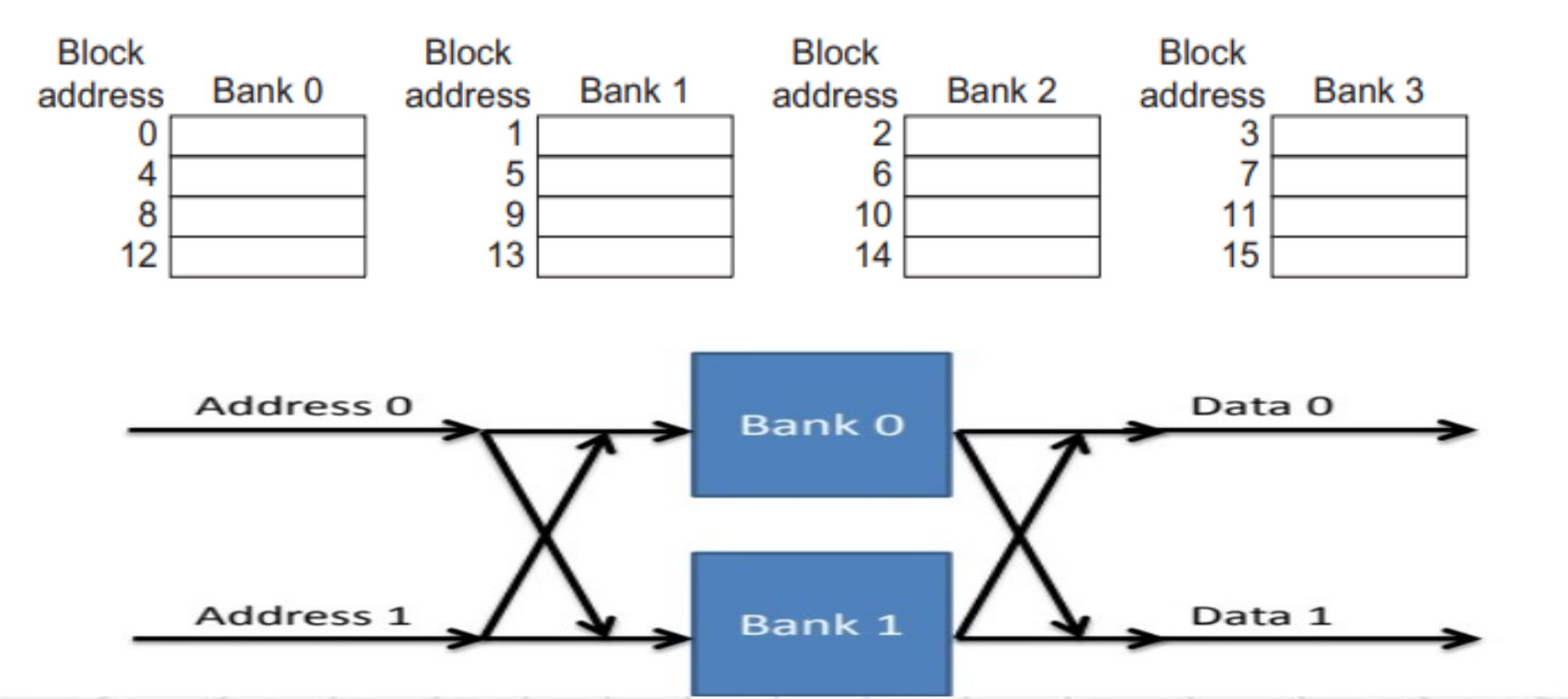
  break a single SDRAM into multi banks;

  they can operate independently;
- Provide some of the advantages of interleaving
- Help with power management

What is Multiple Banks?



What is Multiple Banks?



- 现代缓存系统中，地址的高低位都可以用来选择缓存分区，原因如下：
- 数组（Array）拥有着相同的高位地址，如果此时利用高位地址来选择分区，那么数组中连续的元素都会放在同一个分区，从而损伤了性能
- 结构数组（Array of Struct）中相同的元素拥有着一样的低位地址，如果此时利用低位地址来选择分区，那么结构数组中相同的元素都会放在同一个分区，连续访问时也会造成性能的下降
- 所以，在现代的处理器设计中，分区缓存（Banked Cache）都会有一个路由装置（Router/Crossbar），交替着选择是用地址的高位还是低位，来选取缓存分区，并且确保了每一条流水线都可以访问所有的分区。
-

Why it benefit for cache bandwidth?

- A multi-bank cache just stripes data into multiple memory module, for example, striping one word (4 bytes) across 4 modules. When you read data from multi-bank cache, you can read it in parallel (read one byte from 4 banks at the same time, instead of 4 bytes from one bank sequentially). The result is that the hit time is reduced.

Something about Cache Bandwidth

- 但分区缓存也不是万能的，它为硬件带来了新的挑战：
- 因为需要 Router/Crossbar，所以在设计上会增加难度，系统的功耗和延迟都会增加
- 分区冲突（Bank Conflict）和使用不均匀（Uneven Utilization）的现象依旧存在

Something about Cache Bandwidth

- 近些年L1 Cache与主存储器容量间的比值不但没有缩小而是越来越大。L1 Cache的大小已经很少发生质的变化了，从Pentium的16/32KB L1 Cache到Alder Lake的80KB L1 Cache，Intel用了足足二十多年的时间。在这二十多年中，主储存器容量何止扩大了两千倍。相同的故事也发生在L2与L3 Cache中。

Alder Lake Cache							
AnandTech	Cores P+E/T	L2 Cache	L3 Cache	IGP	Base W	Turbo W	Price $1ku
i9-12900K	8+8/24	8x1.25 2x2.00	30	770	125	241	$589
i9-12900KF	8+8/24	8x1.25 2x2.00	30	-	125	241	$564
i7-12700K	8+4/20	8x1.25 1x2.00	25	770	125	190	$409
i7-12700KF	8+4/20	8x1.25 1x2.00	25	-	125	190	$384
i5-12600K	6+4/20	6x1.25 1x2.00	20	770	125	150	$289
i5-12600KF	6+4/20	6.125 1x200	20	-	125	150	$264

The Pentium III Hits 1 GHz

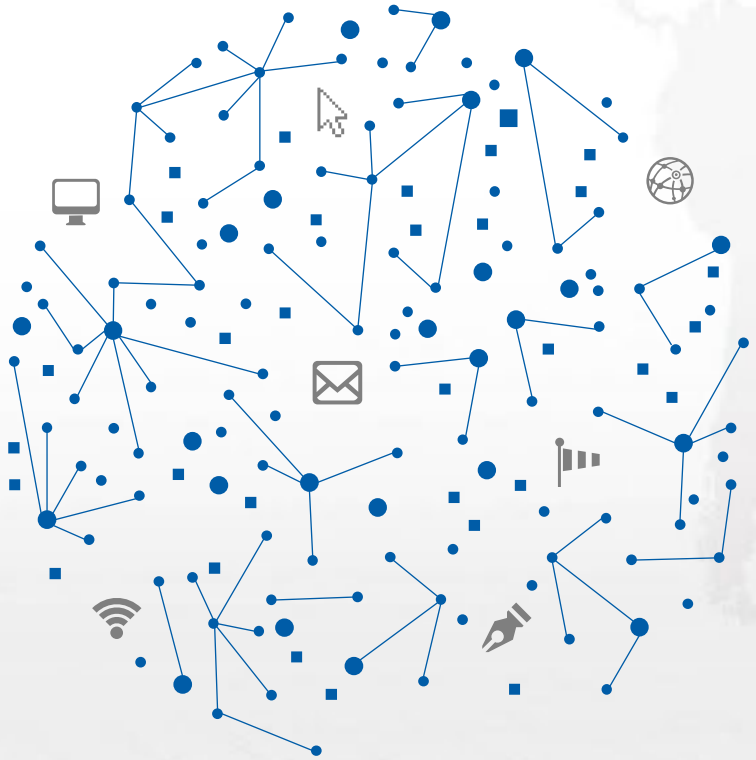
Code name	Coppermine
L1 cache	16 KB + 16 KB
L2 cache	internal, 256 KB (CPU frequency)
Clock frequency	500–1,133 MHz
FSB	100-133 MHz

S R T P 立 项 答 辩

感谢聆听及指导！

浙江大学计算机学院2020级

答辩人：张云策



终于结束了（逃）