

存储系统层次结构

Memory hierarchy



存储器分类



- 按存储介质分
- 按存取方式分
- 按信息的可保存性分
- 按在计算机系统中的作用分



按存储介质分

■ 半导体存储器

- 速度快、功耗低

■ 磁存储器

- 磁芯、磁带、磁盘
- 容量大，速度慢、体积大

■ 激光存储器

- DVD-ROM DVD-RW DVD-R
- 便于携带，廉价，易于保存



按存取方式分

▪ 随机存储器

- 存储器中的任意存储单元都能随机存取
- 且存取时间与物理位置无关
- 磁芯、半导体存储器



▪ 顺序存储器

- 存储器存取时间与物理位置有关
- 磁盘、磁带、激光存储器






按信息的可保存性分



- 易失性存储器 *Volatile Memories*

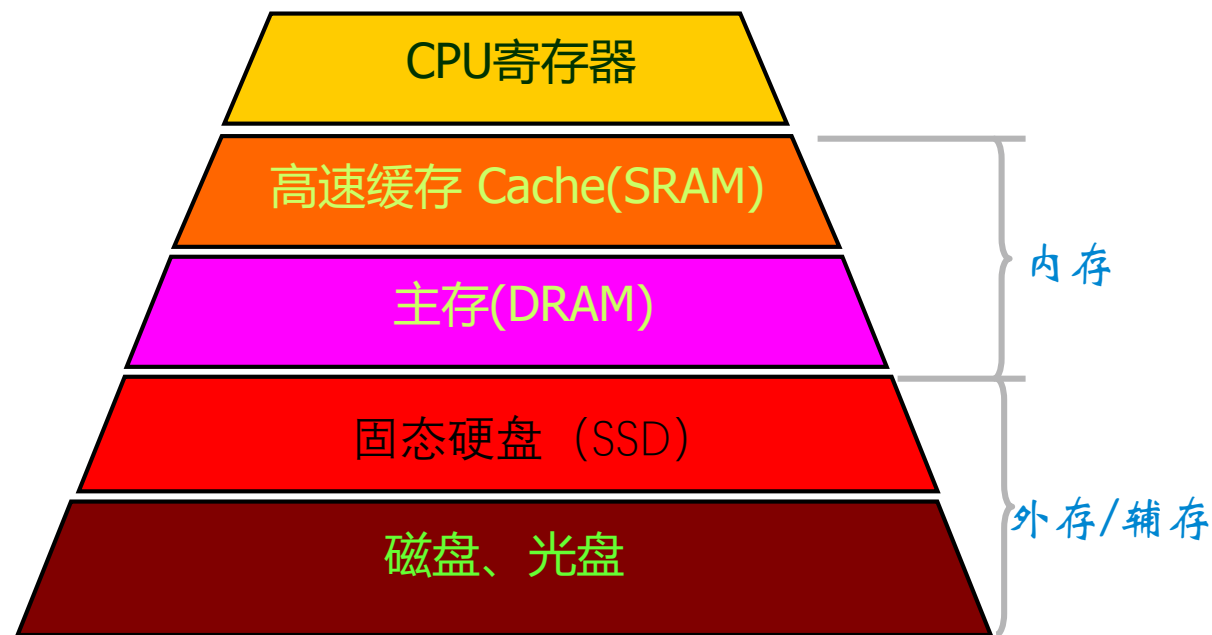
- 断电后信息消失
- SRAM (静态随机访问存储器) 
- DRAM (动态随机访问存储器)

- 非易失性存储器 *Non-Volatile Memories*

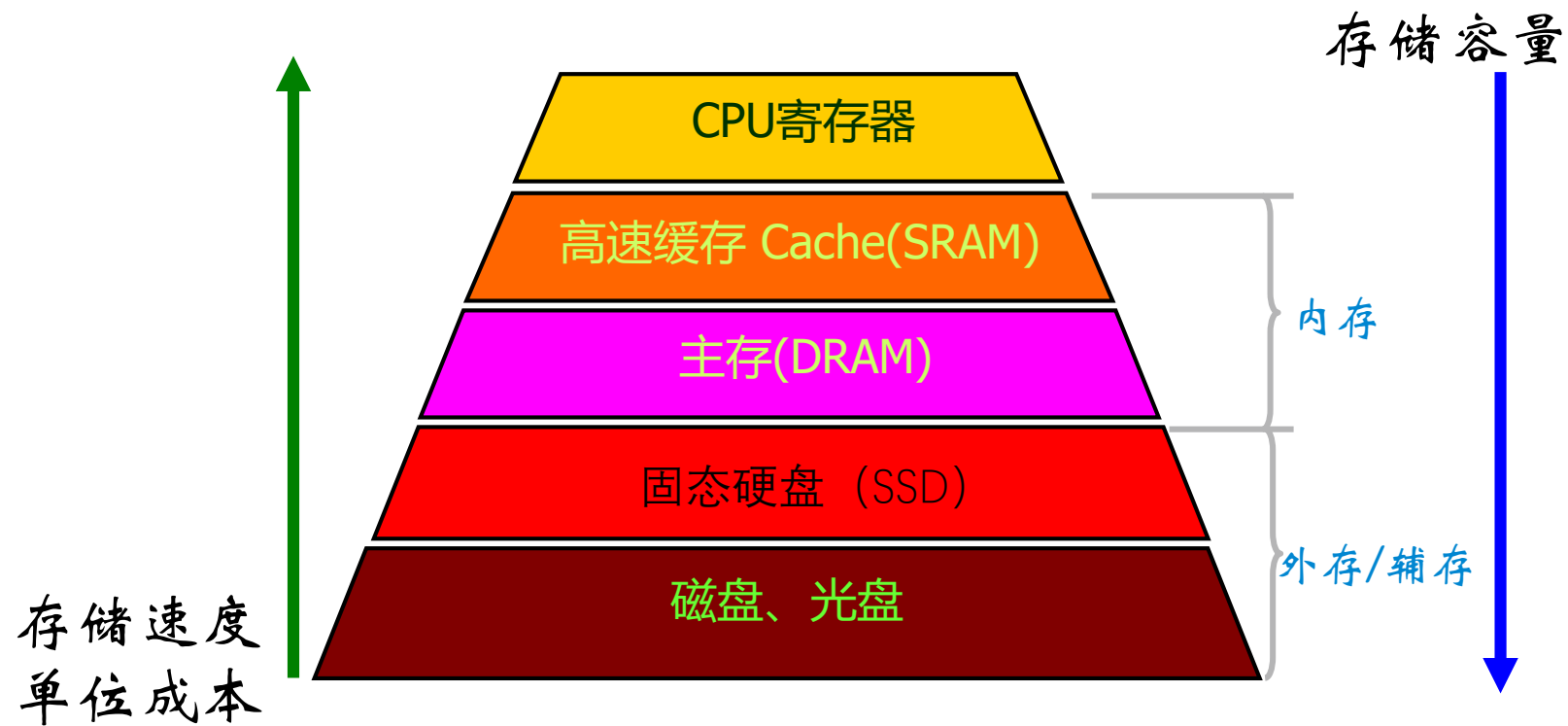
- 断电后仍能保存信息
- 磁存储器、激光存储器、NVRAM (非易失性随机访问存储器)

按在计算机系统中的作用分

- 高速缓冲存储器 Cache
- 主存储器
- 辅助存储器



存储系统分层结构



内存（DRAM）技术的发展趋势

内存：动态随机访问存储器		DRAM
Year	容量	访存周期时间
1980	64 kbit	250 nsec
1983	256 kb	220 ns
1986	1 Mb	190 ns
1989	4 Mb	165 ns
1992	16 Mb	145 ns
1995	64 Mb	120 ns
1999	128 Mb	110 ns
2003	512 Mb	90 ns
2007	2 Gb	60 ns

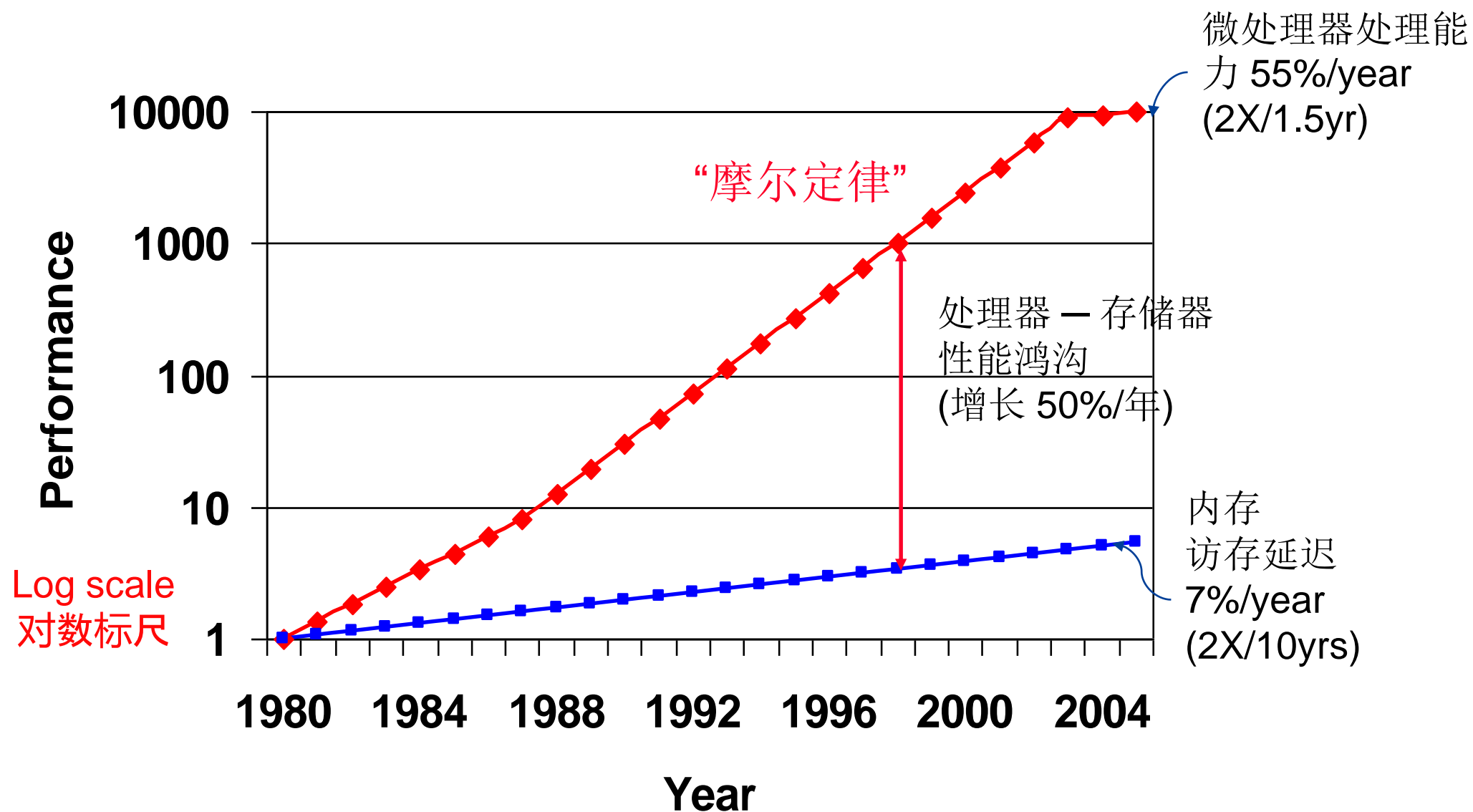
容量 Capacity

DRAM: 4 x in 3 years

访存周期 (延迟: latency)

DRAM: 2 x in 10 years

处理器-存储器性能鸿沟



如何克服性能鸿沟？



程序中（指令和数据）的访存特性：

- ① 程序执行时，除了少部分的转移和过程调用指令外，在大多数情况下仍是顺序执行的。
- ② 过程调用将会使程序的执行轨迹由一部分区域转至另一部分区域，但过程调用的深度在大多数情况下都不超过5。
- ③ 程序中存在许多循环结构，只由少数指令构成，但是它们将多次执行。
- ④ 程序中包括许多对数据结构的处理，如对数组进行操作，往往都局限于很小的范围内。

局部性原理

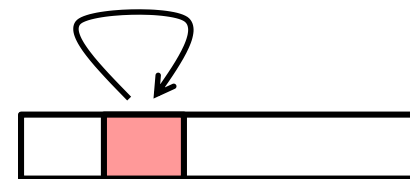


▪ 局部性 (locality)

- 程序倾向于使用最近访问过的相同地址、或相邻地址的数据和指令

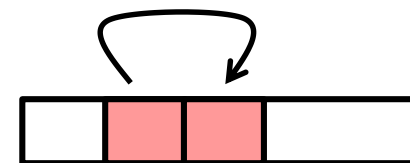
▪ Temporal locality (时间局部性) :

- 最近访问的项目，在不久的将来还会被访问



▪ Spatial locality (空间局部性) :

- 与当前访问项目地址邻近的项目很可能会同时被访问



局部性原理举例



```
sum = 0;
for (i = 0; i < n; i++)
    sum += a[i];
return sum;
```

- 数据访问

- 连续访问数组元素(步长: 1)
- 每次循环都访问变量 i, sum

空间局部性 Spatial locality

时间局部性 Temporal locality

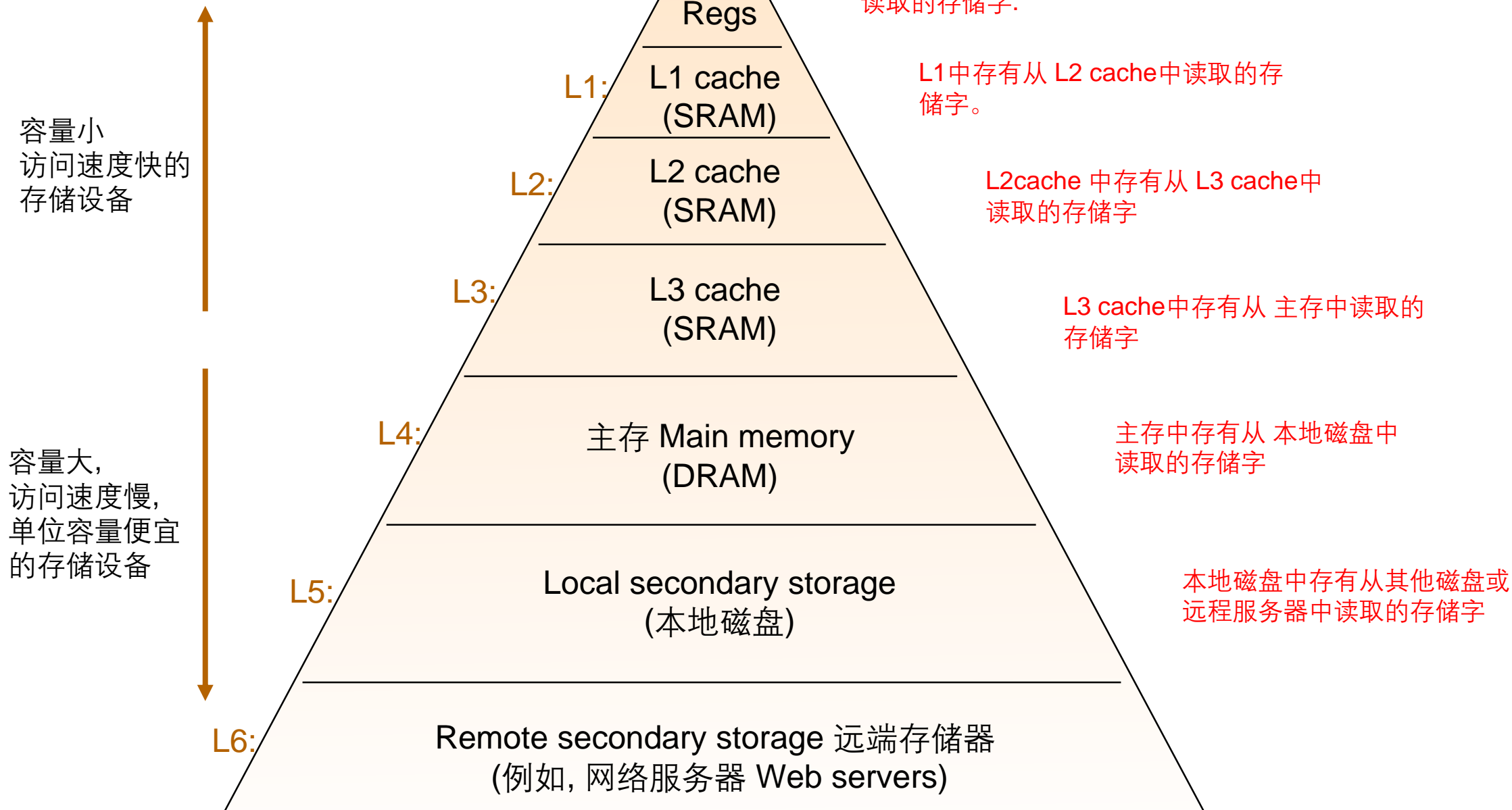
- 指令访问

- 顺序读取指令
- 重复执行循环体内的指令

空间局部性 Spatial locality

时间局部性 Temporal locality

存储系统层次结构



小结



- 存储器的分类
- 处理器-存储器性能鸿沟
- 程序的访存局部性原理
- 存储系统层次结构
- 接下来，介绍各层次中主要部件的工作原理